



Istituto nazionale per la valutazione del sistema educativo di istruzione e di  
formazione

## **WORKING PAPER N. 31/2016**

---

### **La formazione professionale degli insegnanti: quali evidenze abbiamo sulla sua efficacia?**

Giovanni Abbiati – FBK-IRVAPP, [abbiati@irvapp.it](mailto:abbiati@irvapp.it)  
Andrea Caputo – UNIVERSITÀ LA SAPIENZA”, [caputo.and@libero.it](mailto:caputo.and@libero.it)  
Letizia Giampietro – INVALSI, [letizia.giampietro@INVALSI.it](mailto:letizia.giampietro@INVALSI.it)  
Barbara Romano – ASVAPP, [bromano@prova.org](mailto:bromano@prova.org)

*Le opinioni espresse in questo lavoro sono attribuibili esclusivamente agli autori e non impegnano in alcun modo la responsabilità dell'Istituto. Nel citarlo, non è, pertanto, corretto attribuire le argomentazioni ivi espresse all'INVALSI*

## Introduzione

Nel panorama degli studi sull'istruzione, il tema della formazione professionale degli insegnanti ha acquisito negli ultimi vent'anni una rilevanza sempre maggiore. La crescita di questo ambito di ricerca è dato da numerosi fattori, tra i quali la riconosciuta centralità della figura del docente nello sviluppo delle competenze degli allievi (OECD, 2005; Scheerens, 2000) e la conseguente necessità di individuare metodi per renderla più efficace. I risultati di queste ricerche trovano riscontro nell'accresciuta attenzione da parte dei *policy makers* alla gestione del processo formativo dei docenti sia in fase di selezione, sia durante le fasi successive della carriera. Attenzione che si traduce nel riconoscimento della formazione continua come parte essenziale della professionalità docente (Commissione Europea, 2004) e tramite la strutturazione di sistemi di valutazione della qualità della formazione erogata (Eurydice, 2006). Il focus sulla formazione in servizio, in particolare, nasce dalla necessità di fornire ai docenti strumenti *efficaci* per confrontarsi con un contesto mutevole, in cui la padronanza di nuovi linguaggi e nuovi codici assume una crescente importanza (Commissione Europea, 2004; Moscati, 2000). Ai docenti, infatti, si richiede non solo di imparare a trasmettere nozioni, ma anche di imparare a imparare sempre più velocemente (Eurydice, 2000). La necessità di un maggiore coinvolgimento in processi di apprendimento continuo è percepita anche dai docenti, i quali riportano un elevato bisogno di formazione professionale, soprattutto in paesi come l'Italia in cui questa è stata a lungo trascurata (Moscati, 2000; OECD, 2009).

Numerosi studi hanno cercato di stabilire se e a quali condizioni la formazione professionale sia efficace nel migliorare il metodo di insegnamento dei docenti e la sua eventuale ricaduta sugli apprendimenti degli studenti. Tuttavia sono pochi gli studi che utilizzano metodi rigorosi e producono risultati robusti. Inoltre, le lezioni apprese da questi risultati non sono sempre facilmente individuabili, in quanto la forte variabilità tra i diversi programmi di formazione professionale rende difficile stabilire quali siano le caratteristiche chiave per determinarne l'efficacia. Alla luce dell'investimento che è attualmente in essere (e di quello che verrà programmato in futuro) appare quindi di particolare rilievo riuscire a capire come sia possibile convogliare le risorse esistenti verso interventi di formazione realmente capaci di raggiungere gli obiettivi che si sono prefissati.

Lo scopo di questo lavoro è quindi duplice: da un lato fare una rassegna delle esperienze di valutazione degli effetti della formazione professionale e delle diverse metodologie adottate; dall'altro tentare di individuare gli elementi cruciali per una formazione professionale "efficace" selezionando, tra i vari contributi, quelli la cui evidenza può essere considerata abbastanza robusta da mostrare un effetto causale credibile della formazione sull'apprendimento degli studenti e sul

metodo di insegnamento.

La rassegna è quindi così organizzata: nella sezione 1 introduciamo i principali modelli teorici sulla formazione degli insegnanti; nella sezione 2 specifichiamo cosa intendiamo per formazione professionale degli insegnanti e su quale tipo di programmi ci soffermeremo; la sezione 3 passa in rassegna i diversi ambiti di ricerca sulla formazione degli insegnanti e la ricerca valutativa sugli effetti della formazione con particolare riguardo agli studi sperimentali; la sezione 4 invece descrive il lavoro di selezione e analisi dei contributi e discute i risultati della rassegna.

## **1. La formazione degli insegnanti: uno sguardo ai modelli teorici**

Esistono varie definizioni di formazione. In senso generale, per formazione si intende l'acquisizione di competenze legate al ruolo professionale (Villegas-Reimers, 2003). Nello specifico dell'insegnamento, si tratta di una crescita professionale che il docente acquisisce come risultato della propria esperienza lavorativa e a seguito dell'osservazione sistematica del proprio modo di insegnare (Glatthorn, 1995), dove per osservazione si intende un esame personale e critico (svolto in maniera informale) del proprio operato. In alternativa, la formazione degli insegnanti può essere definita come l'insieme di attività volte a sviluppare le competenze e le conoscenze necessarie al docente nel suo lavoro (OECD, 2009).

Da un punto di vista teorico, esistono tre principali *approcci* alla formazione degli insegnanti: l'approccio olistico, l'approccio basato sulla competenza e l'approccio dinamico integrato.

L'*approccio olistico* (Calderhead, 1989; Clift, Houston, & Pugach, 1990) rappresenta l'approccio dominante adottato nei programmi di formazione che sono maggiormente volti a incoraggiare la riflessione sulle pratiche, le esperienze e le credenze dei docenti relative all'insegnamento. In linea con quanto affermato da Elliot (2002), la competenza dei docenti e la misura in cui questi possono migliorare dipende fortemente dalla loro abilità a interrogarsi continuamente sulle condizioni che regolano le proprie interazioni con gli studenti. Particolare enfasi è, quindi, attribuita alle capacità riflessive che concernono l'osservazione, l'analisi, l'interpretazione e il decision-making.

L'*approccio basato sulla competenza* (Thomson, 1991; Whitty & Willmott, 1991) promuove invece lo sviluppo professionale dei docenti attraverso l'acquisizione di abilità didattiche specifiche e ben codificate. Alcuni esperti (Sprinthall, Reiman, & Thies-Sprinthall, 1994) hanno definito un elenco di strategie caratterizzate da un elevato grado di esplicitazione (come rivolgersi agli studenti, come esprimere apprezzamenti sul loro operato, come formulare domande) che i docenti dovrebbero possedere.

Infine, l'*approccio dinamico integrato* (DIA; Antoniou & Kyriakides, 2013) si basa sull'assunto che il miglioramento dell'efficacia dei docenti non possa essere focalizzato sull'acquisizione di competenze o abilità isolate (Gilberts & Lignugaris-Kraft, 1997), né sulla riflessione sull'intero processo di insegnamento senza considerare i bisogni professionali e le priorità di sviluppo dei docenti. Secondo questo approccio ciò che sembra rilevante è la fase evolutiva di sviluppo professionale in cui si colloca il docente; in tal senso, i programmi di formazione dovrebbero essere disegnati in stretta coerenza con i bisogni dei docenti a cui sono rivolti, anche in funzione della loro esperienza professionale.

## **2. Definizioni operative di “Formazione degli insegnanti” e focus della rassegna**

Prima di passare alla rassegna della letteratura è necessario definire cosa si intenda per formazione professionale e a quale accezione faremo riferimento.

Il primo criterio di classificazione è costituito dal *tipo* di formazione professionale degli insegnanti che si intende indagare: l'OECD (2005), infatti, ne distingue tre tipologie che corrispondono a *momenti della carriera* di un docente molto diversi:

- formazione iniziale (pre-service education), che costituisce la formazione universitaria o post-universitaria precedente l'ingresso in servizio;
- accompagnamento (induction programmes), che contempla i programmi di supporto ai docenti nei primi anni dell'ingresso in servizio volti a contrastare il rischio di *dropout* professionale;
- sviluppo professionale (professional development programmes o in-service training programmes), che fa riferimento all'aggiornamento professionale in servizio.

Questa rassegna si focalizza sulla terza tipologia (formazione professionale in servizio, d'ora in avanti formazione) per due ragioni: in primo luogo per la maggiore ricchezza di contributi di ricerca in questo ambito e, secondariamente, per l'immediata utilizzabilità dei risultati nel contesto italiano.

In concreto, lo sviluppo professionale può avvenire sia formalmente, a seguito di esperienze formative strutturate in cui è richiesto l'intervento di un esperto esterno alla scuola (ad esempio nell'ambito di seminari, workshop, conferenze) che informalmente, attraverso un variegato insieme di esperienze formative non strutturate, come gli scambi informali con i colleghi, la lettura di riviste specializzate, la partecipazione a *network* professionali (Ganser, 2000; OECD, 2009). Come mette in evidenza Glatthorn (1995), lo sviluppo professionale così inteso abbraccia quindi una dimensione più ampia dell'avanzamento di carriera o della formazione tenuta in contesti formali, e si riferisce ad ogni tipo di attività che ha come fine l'acquisizione di competenze. Queste definizioni, molto ampie e aperte, ben si adattano al variegato mondo della formazione: come illustra l'OECD (2009) le varie iniziative formative assumono molteplici forme (workshop, lavori di gruppo, conferenze,

percorsi di ricerca-azione) e affrontano svariati argomenti (didattica, esclusione sociale, bullismo, aggiornamento sulle normative).

La letteratura empirica a riguardo, tuttavia, si è concentrata su un sottoinsieme ben definito di iniziative, ovvero gli interventi centrati sul contenuto curricolare e sulla didattica che hanno come obiettivo dichiarato il miglioramento dei risultati degli studenti<sup>1</sup>. La quasi totalità del lavoro di ricerca nel campo della valutazione di singoli interventi è, infatti, limitato a programmi di formazione a) strettamente legati al lavoro di classe con gli studenti e b) relativi a tre materie: lingua nazionale, matematica e scienze. Ciò dipende in larga misura dall'interesse per i programmi che possono avere una maggiore ricaduta pratica nella didattica e che si concentrano sulle discipline "cardine", per le quali esiste in tutti i paesi presi in considerazione ampia disponibilità di test standardizzati attraverso cui misurare l'impatto degli interventi.

All'interno di questo insieme più ristretto di interventi, Kennedy (1998), propone una tipologia che raggruppa i contenuti dei corsi di formazione in quattro grandi famiglie:

1. la prima comprende la formazione sul metodo didattico in generale, mediante l'illustrazione di comportamenti e modalità di insegnamento da adottare in classe utilizzabili indistintamente per tutte le materie scolastiche;
2. la seconda si focalizza sulla didattica disciplinare. I principi di base (es: lavoro di gruppo) di portata generale e comuni ad altre materie, sono declinati per lo specifico della disciplina;
3. la terza famiglia si focalizza sull'insegnamento del curriculum facendo ricorso a strumenti pedagogici, legati cioè ai processi di apprendimento degli studenti;
4. la quarta prevede interventi mirati alla comprensione dei processi di apprendimento e di valutazione della preparazione degli studenti. Non sono interventi legati a particolari pratiche didattiche o comportamenti da adottare in classe, anche quando fanno riferimento a una disciplina specifica.

Come si può notare i quattro tipi individuati da Kennedy possono essere raggruppati in funzione dell'enfasi sui contenuti disciplinari (famiglie 2 e 3 vs famiglie 1 e 4) o sulla didattica (1 e 2 vs 3 e 4). L'autore individua la frattura maggiore tra i gruppi 1 e 2, che si propongono di agire sui comportamenti come mezzo per cambiare i risultati degli studenti, e i gruppi 3 e 4, che invece mirano a fornire conoscenze ai docenti sul modo in cui gli studenti apprendono, senza essere particolarmente prescrittivi sui comportamenti da adottare. Seguendo la numerazione proposta dall'autore, i gruppi possono essere collocati su un continuum che li ordina dal più al meno

---

<sup>1</sup> Anche altri tipi di interventi non direttamente legati all'insegnamento, a onor del vero, possono avere ricadute positive sul lavoro di classe. Si pensi ad esempio a corsi sulla multiculturalità o sull'alfabetizzazione informatica che, avvicinando gli insegnanti alle esigenze degli alunni, possono contribuire in maniera indiretta a migliorare il clima di classe e di conseguenza i processi di apprendimento.

prescrittivo (ibidem).

Rimane ancora una notevole variabilità rispetto all'implementazione. In questo caso però non sono state sviluppate delle tipologie vere e proprie, ma sono state piuttosto individuate alcune caratteristiche che sono considerate potenzialmente utili nel garantire una implementazione efficace. A tale riguardo, Garet e collaboratori (2001) individuano tre caratteristiche di base nel disegno dei programmi, di seguito descritte:

- focus sui contenuti specifici, in cui la preferenza è accordata a programmi focalizzati su contenuti specifici (disciplinari o meno) rispetto a quelli che affrontano tematiche generali;
- opportunità di apprendimento attivo, in cui i programmi più efficaci sono quelli che permetterebbero ai docenti di sperimentare quanto appreso in classe e monitorare i feedback propri e degli studenti a quanto appreso nella formazione e successivamente praticato;
- coerenza con altre attività di formazione, dal momento che affinché il programma sia percepito importante e meritevole di impegno dagli insegnanti, occorre che questo sia inserito all'interno di una chiara strategia istituzionale che ne indichi la priorità.

Oltre a queste, sono state individuate ulteriori tre caratteristiche di tipo strutturale:

- tipo di attività: sono considerate esclusivamente attività basate su gruppi di lavoro, attivazione di reti di scuole, forme di mentoring e coaching. In altri termini, le tradizionali conferenze e i seminari di un giorno non vengono considerati come potenzialmente efficaci.
- durata: preferibilmente elevata e strutturata nel corso di più incontri, che siano distribuiti lungo l'intero anno scolastico.
- partecipazione collettiva: riprendendo la letteratura sull'educational effectiveness, un ultimo ingrediente sembra essere costituito dalla collaborazione tra gruppi di docenti della stessa scuola, perché possano interagire e aiutarsi sui contenuti dei corsi o tramite l'attivazione di processi di apprendimento informale.

Restringere il campo di ricerca attorno a questo tipo di caratteristiche di contenuto, di disegno e di struttura raccoglie un certo consenso nel mondo della ricerca, che raramente ha valutato altri tipi di programmi, soprattutto quelli di durata molto breve o basati su sporadiche iniziative quali convegni o seminari. L'evidenza alla base di queste raccomandazioni, tuttavia, non è molto robusta: le caratteristiche sopraelencate sono state individuate dal gruppo di ricerca coordinato da Garet tramite un'analisi campionaria dei programmi percepiti come efficaci dagli insegnanti. Tuttavia, esiste un'ampia convergenza circa la loro importanza, come sembrano segnalare altri studi simili (Desimone et al., 2002; Phillips et al., 2011; Sailors & Price, 2010). In particolare, i ricercatori condividono un'esplicita avversione per la prassi consolidata dei corsi generalisti della durata di un giorno, i quali costituiscono ancora la forma preponderante (Garet et al., 2001; Sailors & Price,

2010), oltre che la meno apprezzata dai docenti: oltre il 90% degli insegnanti statunitensi interpellati sul tema trova questo tipo di formazione noiosa e inutile, oltre ad ammettere di dimenticarsi immediatamente dei contenuti (Miller, 1998). Si potrebbe a questo punto pensare che includere tali caratteristiche nei programmi di formazione renda superflua una valutazione della loro efficacia o, più prudentemente, che gli esiti della valutazione di programmi così disegnati siano sostanzialmente convergenti. Al contrario, le ricerche empiriche che mostrano evidenza solida sugli effetti della formazione professionale registrano ampie divergenze sia in quanto alla dimensione (o addirittura presenza) dell'effetto, sia rispetto alle caratteristiche degli interventi che sembrano produrlo.

### **3. La ricerca sulla formazione professionale**

In letteratura è possibile distinguere tre principali ambiti di ricerca nell'area della formazione professionale degli insegnanti.

Un primo ambito di studi, di matrice prevalentemente pedagogica e psicosociale, si occupa della ricerca qualitativa connessa ai processi che vengono messi in atto durante il momento formativo, per lo più attraverso metodi come l'etnografia, la *grounded theory* e gli studi di caso. Le tecniche utilizzate spaziano dall'osservazione, alle interviste, ai focus group, con differenti livelli di strutturazione. Questo tipo di studi non intende indagare legami di natura causale tra la formazione e un definito insieme di *outcome*, quanto piuttosto fornire un'analisi descrittiva e analitica delle variabili che più concorrono al processo di apprendimento da parte dei docenti dei contenuti della formazione. A tale riguardo, diversi studi evidenziano come l'apprendimento sia realizzabile attraverso un cambiamento di opinioni e atteggiamenti rispetto all'insegnamento (Eilks, 2011), lo sviluppo di rappresentazioni stabili relative ai contenuti della formazione (Franke, 2001) e una maggiore disponibilità ad adottare e applicare tali contenuti nella pratica quotidiana (Jimoyiannis 2010; Roehrig 2011).

Un secondo ambito di studi concerne la ricerca correlazionale tesa a esplorare il possibile rapporto tra la formazione degli insegnanti e i risultati degli studenti. Si tratta di studi basati su campioni molto ampi in cui sono presenti informazioni sul rendimento degli studenti agganciate a dati anagrafici degli insegnanti e caratteristiche connesse al loro livello di formazione e aggiornamento professionale. In questi studi non si tratta quindi di valutare propriamente l'efficacia di un singolo corso di formazione ma la partecipazione *a un qualche tipo di formazione*, a prescindere dalle sue caratteristiche. La formazione del docente diviene una "proprietà" del docente stesso: non a caso, lo scopo della maggior parte di questi studi è comprendere quali caratteristiche

del docente siano collegate all'apprendimento dei suoi studenti. Wayne e Youngs (2003), mettendo a rassegna questo tipo di letteratura, rilevano che – generalmente – vi è una associazione tra i docenti che hanno ricevuto una formazione sul contenuto disciplinare e sulla didattica e performance più elevate degli studenti, soprattutto nel caso dei docenti di matematica. Questi risultati sono stati recentemente corroborati anche da Clotfelter et al. (2007). Occorre rimarcare, tuttavia, che questi risultati non riescono a risolvere un problema fondamentale: i dati utilizzati, infatti, essendo di natura osservazionale, non permettono di distinguere chiaramente tra fenomeni di autoselezione dei docenti all'interno dei corsi di formazione ed effetto causale della stessa. In altre parole, non è possibile distinguere se siano i docenti migliori coloro che si impegnano in corsi di aggiornamento post laurea e che hanno frequentato le università migliori, oppure se siano i corsi e le università che preparano un docente a essere più efficace nell'insegnamento.

Il terzo filone di studi utilizza metodi sperimentali e non sperimentali per stimare l'impatto della formazione. Recentemente, studi di tipo controfattuale, ma non sperimentali, hanno utilizzato discontinuità temporali nell'erogazione di corsi di formazione (Angrist & Lavy, 2001; Jacob & Lefgren, 2004) o tecniche di abbinamento statistico (Correnti, 2007) per risolvere il problema dell'attribuzione causale. Benché l'evidenza raccolta da questi studi sia abbastanza robusta, i risultati prodotti sono difficilmente utilizzabili nella prassi concreta. Questi studi, infatti, indagano l'efficacia di vasti programmi di formazione a livello nazionale o locale, nei quali confluiscono una molteplicità di iniziative eterogenee. Jacob e Lefgren (2004) ad esempio, trovano un effetto nullo della formazione, a differenza di Angrist e Lavy (2001), i quali invece rilevano un effetto positivo (anche se non sempre significativo) per i docenti appartenenti a scuole non religiose. La difficoltà nell'estrarre indicazioni di *policy*, inoltre, è acuita dalla mancanza di una descrizione delle caratteristiche dei programmi (e della loro implementazione) da parte degli autori. Come abbiamo visto, infatti, la variegata gamma di corsi di formazione offerti ai docenti non è sempre riconducibile direttamente al lavoro di classe (ad esempio, corsi di aggiornamento sulle tecnologie informatiche o sulla normativa scolastica) oppure è costituita da attività frammentate e di durata molto breve.

Per questa ragione nelle successive sezioni di questo lavoro ci concentreremo esclusivamente su studi di tipo sperimentale. Due elementi fondamentali li differenziano dai precedenti. Il primo è la focalizzazione su *uno specifico programma di formazione*, con una precisa descrizione delle sue caratteristiche e delle modalità di implementazione. La ricchezza di informazioni contenute in questi studi, quindi, va di pari passo con la ristrettezza del campo a cui si applicano le conclusioni. Il secondo è l'utilizzo del metodo sperimentale, che, grazie all'assegnazione casuale della partecipazione al programma tra i soggetti ammissibili (e la conseguente creazione di un gruppo di

trattamento e un gruppo di controllo), consente di stimare gli effetti confrontando l'*outcome* post-intervento per i due gruppi che, proprio in virtù della randomizzazione, sono equivalenti in tutto tranne che per l'esposizione al programma. Alla luce di quanto visto finora, la scelta di confinare la nostra analisi a questo tipo di letteratura è motivata da un doppio ordine di motivazioni, di natura metodologica e sostantiva. Dal punto di vista del metodo, gli studi di natura sperimentale possiedono gli standard più rigorosi, i quali permettono verosimilmente di attribuire le differenze riscontrate su un insieme di *outcome* di interesse tra i due (o più) gruppi di soggetti al programma stesso distinguendolo da eventuali componenti spurie<sup>2</sup>. Dal punto di vista sostantivo, invece, l'attenzione concentrata su un singolo programma permette di avere una conoscenza più approfondita delle sue caratteristiche e di metterle, quindi, in relazione agli effetti, consentendo quindi l'individuazione di *caratteristiche chiave* per una formazione professionale efficace.

Le rassegne di contributi sull'argomento (e.g. Yoon, 2007) mostrano risultati incoraggianti: numerosi programmi di formazione progettati secondo i criteri proposti da Garet et al. (2001) impattano positivamente sull'apprendimento. Questo risultato sembra esser vero nelle tre materie principali (scienze, lingua nazionale, matematica). Tuttavia, accanto a programmi efficaci ne compaiono altrettanti privi di effetti misurabili sugli studenti (per un esempio si veda Garet et al., 2011). Il fatto che programmi in possesso delle caratteristiche che li annoverano tra quelli potenzialmente efficaci abbiano mostrato effetti così diversi impone di ricercare più a fondo quali siano le proprietà fondamentali che un corso dovrebbe possedere. Al momento, queste proprietà non sono ben note (Guskey, 2009), anche a causa dell'esiguità dei contributi sperimentali inclusi fino ad oggi nelle rassegne (solo 9, ad esempio, in Yoon, 2007). Negli ultimi cinque anni, tuttavia, una crescente mole di evidenza sperimentale consente, forse, di fare qualche progresso nell'identificazione delle caratteristiche irrinunciabili dei corsi di formazione.

#### **4. L'efficacia della Formazione degli insegnanti secondo gli studi sperimentali**

Torniamo alle domande alle quali ci siamo proposti di rispondere con questa rassegna:

- la formazione degli insegnanti è efficace nel migliorare le loro conoscenze e abilità?
- la formazione degli insegnanti è efficace per migliorare l'apprendimento degli studenti?
- c'è evidenza su quali siano i meccanismi che rendono la formazione degli insegnanti efficace?

---

<sup>2</sup> Ricordiamo che anche il metodo sperimentale non è immune da problemi. Berk (2005) a questo proposito conia il termine *bronze standard*, per definire gli esperimenti randomizzati ed evidenziare gli aspetti problematici di un metodo che era da sempre considerato il *gold standard* della ricerca valutativa. L'autore, nel fare ciò, comunque riconosce che "*randomized field experiments are still the best way to estimate causal effects, but are a considerable distance from perfection*" (*ibidem*, p.1).

L'esaustiva rassegna della letteratura sulla formazione proposta da Yoon (2007) copre tutti gli studi realizzati fino al 2005 e rappresenta il nostro punto di partenza. Tale rassegna si è focalizzata esclusivamente sugli effetti prodotti sugli studenti, mentre in questa sede intendiamo considerare gli effetti sia a livello studente che a livello insegnante, in quanto occorre essere sicuri che la formazione sia in primo luogo efficace per gli insegnanti stessi: solo così ha *chances* di essere efficace anche per gli studenti. Se si osservano esclusivamente gli effetti sugli studenti e non si rileva alcun effetto sugli insegnanti, diventa impossibile inferire i meccanismi causali in grado di garantire un efficace trasferimento di competenze tra docenti e studenti.

Di seguito, illustriamo il processo di selezione degli studi che, all'interno dell'ampia produzione di ricerca sulla formazione degli insegnanti, abbiamo utilizzato per le nostre analisi al fine di dare risposta alle domande di ricerca.

#### **4.1 Approccio metodologico**

##### ***La selezione degli studi potenzialmente includibili***

Al fine di selezionare i contributi da analizzare è stata effettuata una revisione sistematica della letteratura internazionale, ossia un lavoro di analisi, sintesi e valutazione critica dei principali contributi pubblicati. Il lavoro è stato condotto utilizzando le banche dati della produzione scientifica internazionale indicizzata e alcuni dei più importanti motori di ricerca in tema di educazione e formazione: IES, ERIC, WWCH, JSTOR, Best evidence encyclopedia, IZA, Opengrey.eu, OECD, Sociological abstracts, Campbell Collaboration Library, Web of knowledge, PsychInfo. In aggiunta, sono state consultate alcune delle riviste più importanti nell'ambito della ricerca e della valutazione educativa, tra cui l'American Educational Research Journal, l'Educational Evaluation and Policy Analysis, la Review of Educational Research e Sociology of Education. Come già premesso, dal momento che il focus della rassegna è sulla formazione in servizio degli insegnanti, sono stati utilizzati come chiave di ricerca i termini "*teacher*" & "*professional development*", "*in-service training*", "*continuing education*", "*life-long learning*", nei campi del titolo e/o dell'*abstract*, limitando la raccolta ai contributi pubblicati dal 2000 al 2012. Dei record prodotti, in seguito a una analisi preliminare condotta sugli *abstract*, sono stati selezionati soltanto i contributi che mantenevano una sufficiente coerenza con l'argomento in esame, scartando quelli in cui le chiavi di ricerca utilizzate erano soltanto citate o riportate in elenco. In dettaglio, sono stati adottati i seguenti criteri di inclusione:

- il focus del contributo è sulla formazione in servizio degli insegnanti;
- per le ricerche empiriche, il contributo fornisce evidenze quantitative relative alla valutazione di un programma/intervento di formazione quale studio dimostrativo o politica a scala.

Al fine di accrescere la nostra capacità di mettere in relazione risultati provenienti da studi diversi e individuare ulteriori chiavi interpretative, sono state consultate anche ulteriori tipologie di contributo, quali articoli qualitativi, articoli teorici o review/meta-analisi.

Ne è risultato così un campione di 129 lavori, di cui 94 articoli di rivista, 13 rapporti tecnici di ricerca, 13 rassegne di esperienze, 5 capitoli su volume, 2 working papers, 1 tesi di dottorato e 1 conference paper. Inoltre, di questi 129, 11 riguardavano contributi teorici, 21 lavori di review e meta-analisi, 16 ricerche di stampo qualitativo, 7 ricerche controfattuali non sperimentali, 31 studi di tipo sperimentale e 43 ulteriori studi di natura quantitativa non sperimentale<sup>3</sup>.

### ***I criteri di inclusione degli studi nella rassegna***

La seconda fase ha previsto una ulteriore selezione degli articoli precedentemente raccolti. Abbiamo esaminato gli abstract di questi studi, eliminando quelli non conformi ai criteri di seguito riportati, in linea con la precedente rassegna di Yoon (2007):

- tipo di documento: contributo di ricerca;
- età degli alunni: i docenti coinvolti negli interventi di formazione devono insegnare nelle scuole dell'infanzia, primaria e secondaria (è esclusa l'educazione post-secondaria);
- materia: i docenti inclusi nella ricerca insegnano matematica, scienze, o lingua nazionale;
- periodo: le ricerche devono essere pubblicate posteriormente al 2005 per evitare sovrapposizioni con la rassegna di Yoon (2007):
- metodo utilizzato: esperimento randomizzato controllato (RCT) o disegno quasi-sperimentale (QED)
- outcome: lo studio deve misurare l'impatto sugli studenti (apprendimento) o sui docenti (conoscenza della materia, atteggiamenti, pratiche didattiche);
- descrizione del PD: il contributo deve riportare una descrizione dettagliata del programma valutato.

Per accrescere la nostra capacità di esaminare l'efficacia della formazione degli insegnanti, abbiamo unito i contributi selezionati in base ai criteri di sopra specificati ai contributi presenti nella rassegna di Yoon (2007). Tra i 129 contributi di ricerca raccolti nella prima fase, ci siamo quindi concentrati su 31 lavori che hanno pienamente soddisfatto tutti i criteri di inclusione.

### ***La codifica degli studi inclusi nella rassegna***

La fase successiva nel processo di rassegna è stata la codifica degli studi che rispondevano a tutti i criteri di inclusione proposti. Abbiamo quindi predisposto tre griglie per registrare le

---

<sup>3</sup> Cfr. Appendice A.

medesime caratteristiche chiave presenti in tutti gli studi analizzati. La prima griglia (Tabella 1. Effetti<sup>4</sup>) fa riferimento agli effetti degli studi rilevati sia sugli insegnanti che sugli studenti: essa include gli *outcome* misurati, l'effetto grezzo (là dove disponibile), la misura dell'effetto espressa in deviazioni standard, la correzione per il clustering. Coerentemente con il framework di *What Works Clearinghouse* lo stesso adottato da Yoon (2007), quando l'effetto non è statisticamente significativo ma l'*effect size* maggiore di 0,25, esso è stato riportato come “statisticamente non significativo ma sostanzialmente rilevante”.

La seconda griglia (Tabella 2. Caratteristiche<sup>5</sup>) riporta le caratteristiche dei programmi di formazione: la disciplina oggetto della formazione (matematica, scienze, lingua nazionale); la popolazione target (nessun target specifico, scuole con scarse performance, scuole con basso reddito, scuole con elevata percentuale di studenti stranieri); il livello dell'intervento (studente o insegnante); il livello scolastico; il contenuto della formazione (breve descrizione degli obiettivi e delle modalità di implementazione), la tipologia di formazione (workshop, corso, coaching); la classificazione secondo le quattro tipologie di 'professional development' proposte da Kennedy; l'intensità (durata prevista del corso in ore di formazione); la partecipazione (ore di formazione svolte sul totale o percentuale di docenti che hanno completato la formazione).

La terza griglia (Tabella 3. Validità<sup>6</sup>) illustra le caratteristiche dello studio: metodologia impiegata per la valutazione degli effetti; numero di distretti scolastici, scuole, insegnanti e studenti coinvolti; livello di *attrition*, ovvero la progressiva perdita o abbandono degli iscritti alla formazione; il livello di randomizzazione (nel caso di RCT) e alcune considerazioni circa la validità interna ed esterna dello studio.

## 4.2 Gli effetti sugli studenti e sugli insegnanti

Gli effetti degli studi sono riportati in termini di *effect size* al fine di consentire il confronto tra studi che utilizzano come misure di *outcome* principalmente punteggi a test espressi con metriche differenti. L'*effect size* si ottiene dividendo l'effetto espresso nella metrica originaria per la deviazione standard dell'*outcome* osservata per trattati e controlli<sup>7</sup>.

### 4.2.1 Effetti sui docenti

#### Perché rilevare l'effetto sui docenti

---

<sup>4</sup> Cfr. Appendice B.

<sup>5</sup> Cfr. Appendice B.

<sup>6</sup> Cfr. Appendice B.

<sup>7</sup> Questa misura è detta Cohen's *d*. Alcuni studi ricorrono, invece al Glass  $\Delta$  che prevede di inserire a denominatore solo la varianza dei controlli, perché quella dei trattati potrebbe essere stata influenzata dal trattamento. In ogni caso le differenze derivanti dall'utilizzo di un metodo o dell'altro sono trascurabili.

La valutazione degli effetti della formazione sui docenti rappresenta un campo relativamente nuovo e poco esplorato: nonostante i destinatari della formazione siano i docenti, questi sono stati spesso visti come dei semplici “tramite” per il passaggio delle competenze agli studenti. Un altro motivo per il ritardo con cui gli effetti sui docenti sono stati valutati, inoltre, deriva dal numero limitatissimo di docenti inclusi nella maggior parte degli esperimenti che ha reso, di fatto, difficile se non impossibile stabilire un collegamento tra la formazione e qualsiasi tipo di *outcome* sui docenti. Recentemente, invece, grazie a un accresciuto interesse per i processi che vengono attivati nel corso della formazione e al maggior numero di soggetti inclusi negli studi di valutazione, è possibile estendere agli insegnanti la rilevazione degli effetti degli interventi formativi. L’interesse per questo tipo di risultati nasce dalla considerazione fondamentale che qualsiasi cambiamento osservabile sugli studenti passi necessariamente attraverso un precedente effetto sui docenti. Nelle parole di Desimone (2009), il docente è il mediatore necessario tra la formazione professionale e i risultati degli studenti.

La letteratura nel campo delle risorse umane individua due modi tramite i quali questo tipo di mediazione può avvenire: generalizzazione e mantenimento (Blume, Ford, Baldwin, & Huang, 2010). Si ha generalizzazione quando quanto appreso in un contesto formativo viene trasferito in contesti e situazioni diverse. Nel caso dell’insegnamento questo si verifica nel passaggio tra formazione ricevuta e lavoro d’aula mediante un’accresciuta conoscenza della materia (importante soprattutto nei campi che richiedono continuo aggiornamento e in quelli che vedono un alto numero di insegnanti cosiddetti “*out-of-field teachers*”, ossia laureati in una materia diversa da quella che insegnano) oppure tramite una maggiore consapevolezza del metodo di insegnamento. Per mantenimento, invece, si intende la misura in cui i cambiamenti prodotti da una esperienza di apprendimento persistono nel tempo. Nel campo dell’insegnamento questo è particolarmente rilevante in quanto l’insegnante formato potrà beneficiare dei contenuti e dei metodi appresi per il resto della sua carriera (che in alcuni paesi – tra cui il nostro – sono estremamente lunghe) e, quindi, potrà con la sua azione didattica raggiungere un numero di studenti molto elevato: i potenziali effetti della formazione, quindi, riguardano un numero di studenti considerevolmente maggiore di quelli inseriti negli studi di valutazione.

Normalmente, le valutazioni dell’efficacia della formazione sui docenti sono accessorie rispetto agli effetti misurati sugli studenti: è raro infatti trovare studi che si concentrino esclusivamente sui docenti (si vedano come eccezioni Grigg 2012; Masters et al., 2010). Non bisogna dimenticare, infatti, che l’*outcome* principale delle ricerche valutative (e il fine ultimo per il quale viene promossa la formazione) è costituito dall’apprendimento degli studenti. A conferma di ciò, le analisi di potenza statistica utilizzate per stabilire il campione minimo da includere nella

ricerca sono calcolate esclusivamente tenendo conto dei risultati attesi sugli studenti: le analisi sui docenti, mancanti di questo pre-requisito, vengono spesso denominate “esplorative”, e i risultati non sempre vengono commentati, anche quando riportati.

### Dimensioni su cui rilevare gli effetti sui docenti

Esistono due dimensioni principali che vengono indagate: la conoscenza della materia e l'adozione di comportamenti/atteggiamenti ispirati dai programmi proposti. Mentre sulla conoscenza della materia i dati sono facilmente comparabili (grazie all'utilizzo di test standardizzati), il panorama degli effetti sulle pratiche didattiche è più sfaccettato in quanto non esiste un *outcome* definito e su cui ci sia un'ampia convergenza in letteratura. In aggiunta, il modo in cui gli effetti sono rilevati è molto eterogeneo: in alcuni studi vengono usati degli osservatori esterni che assistono al lavoro di classe (con procedure e protocolli standardizzati), mentre in altri si ricorre alle auto-dichiarazioni degli insegnanti. Elemento comune a questi approcci è la direzione attesa dell'effetto: i corsi si propongono di modificare le pratiche didattiche dei docenti portandole da una forma di insegnamento tradizionale, basata su lezioni frontali poco interattive, a una *inquiry-based*, basata cioè su metodi che coinvolgano attivamente gli studenti e ne stimolino l'esplorazione, tenendo conto dei risultati di ricerca in campo educativo (Correnti, 2007). Da questo punto di vista sono da considerarsi più robusti gli studi che prevedono un osservatore esterno e sistemi di codifica standardizzati del lavoro di classe<sup>8</sup>. Infatti, la presenza di un protocollo di sperimentazione a cui i docenti dovrebbero attenersi nel lavoro in classe non necessariamente si traduce nell'effettiva attuazione dei comportamenti o delle pratiche didattiche su cui la formazione intende intervenire.

Poco più della metà degli studi esaminati - 17 su 31 - (cfr. Tabella 1) rilevano gli effetti sugli insegnanti. Come accennato nella sezione precedente, esistono alcuni rilevanti problemi metodologici nel comparare direttamente i risultati di questi studi: mentre per gli studenti l'*outcome* è sempre (o quasi) un test standardizzato, per gli insegnanti gli effetti vengono rilevati su dimensioni non assimilabili tra loro e misurate in modi molto diversi:

- a) misure di conoscenza della materia del docente (valutata tramite test standardizzati, come per gli studenti); gli effetti rilevati sulla conoscenza della materia sono espressi sotto forma di *effect size* e quindi facilmente comparabili;
- b) misure di pratiche didattiche, rilevate direttamente da osservatori: anche se spesso viene riportato l'*effect size*, non sempre le dimensioni su cui viene misurato l'effetto sono comparabili;

---

<sup>8</sup> I codificatori, è bene precisare, non conoscono il gruppo di assegnazione dell'insegnante (trattato/controllo).

c) misure di atteggiamento o di pratiche didattiche, rilevate tramite interviste strutturate agli insegnanti; si tratta quindi di misure auto-dichiarate dagli insegnanti relative alle dimensioni più varie, quasi sempre prive di scale validate o consensualmente riconosciute. In questo caso la comparabilità è minima.

La grande eterogeneità di misure e di dimensioni rilevate nei singoli studi impedisce di stabilire un *effect size* medio della formazione professionale (se non per alcuni casi, come vedremo), ma permette comunque di ragionare in termini di presenza/assenza di effetto. Gli effetti della formazione espressi in forma di *effect size*, inoltre, sono evidenziati in 12 studi su 17.

Questa la composizione dei 17 studi di cui discuteremo gli effetti: 8 studi riguardano la lingua nazionale, 5 matematica e 4 scienze; in 6 casi si tratta di interventi per docenti di scuola media, in 8 casi di interventi per le scuole elementari, in 2 casi di scuole superiori e 1 caso tratta di interventi organizzati per elementari e medie.

### Risultati generali sui docenti

I 17 studi che rilevano *outcome* sui docenti, riportano un centinaio di effetti diversi, che risultano statisticamente significativi o sostanzialmente rilevanti in 50 casi.

Dato che numerosi studi riportano i risultati disaggregati per sotto-scale o scompongono gli effetti di una dimensione in più misure, aggregiamo gli oltre 100 risultati trovati per studio e per tipo di dimensione indagata. Ad esempio, in Gersten et al. (2010) sono riportati gli effetti su due dimensioni di pratica didattica e su due dimensioni di conoscenza del docente: in questo caso abbiamo prodotto due medie, una per la dimensione conoscenza della materia e una per la pratica didattica. Chiaramente, l'operazione di calcolo della media è stata effettuata solo in presenza di risultati espressi come *effect size*<sup>9</sup>. Data l'aggregazione effettuata, interpreteremo i risultati alla luce della grandezza dell'*effect size* e non della significatività statistica (elemento che si perde effettuando questa operazione non avendo disponibili informazioni sulla dimensione dei gruppi dei trattati e dei controlli). Complessivamente, 13 programmi sono risultati efficaci in almeno una delle dimensioni rilevate, mentre 4 non sono stati efficaci.

È purtroppo difficile individuare delle linee generali dato l'esiguo numero di studi e la limitata comparabilità degli *outcome*. Nei paragrafi seguenti sono descritti in dettaglio i risultati divisi per tipo di dimensione rilevata (conoscenza della materia, pratiche didattiche osservate, atteggiamenti/comportamenti dichiarati dai docenti).

---

<sup>9</sup> Nei casi degli studi che presentano disegni fattoriali, con più iniziative di PD concorrenti tra di loro, abbiamo riportato i risultati separatamente, trattandoli come studi diversi.

### Conoscenza della materia

Gli studi che si occupano di misurare la conoscenza della materia dei docenti dopo l'esposizione alla formazione sono 5 e riguardano la lingua nazionale (Garet et al., 2008; Masters et al., 2010; Gersten et al., 2010), le scienze (Heller et al., 2012) e la matematica (Santagata, 2011). In 4 studi su 5 viene riscontrato un effetto positivo della formazione. L'*effect size* medio di questi studi si attesta a 0,28, variando da un minimo di 0,05 a un massimo di 0,53. L'unico studio in cui non sono stati rilevati effetti significativi è anche l'unico che riguarda la matematica (*effect size* di 0,1); in lingua nazionale l'*effect size* medio è di 0,36 mentre per scienze è pari a 0,5. Anche a causa del ridotto numero di contributi, non emergono particolari indicazioni circa elementi qualificanti l'intervento che portino a una maggiore efficacia: gli interventi analizzati sembrano quindi essere mediamente efficaci, per lo meno nel breve periodo.

Nei due casi in cui sono stati osservati anche effetti di medio periodo, si osserva che gli effetti trovati alla fine del primo anno di sperimentazione svaniscono (Garet et al., 2008) e che gli effetti rilevati direttamente l'anno successivo alla sperimentazione sono nulli (Garet et al., 2012).

### Pratiche didattiche

Sono 8 gli studi che riportano effetti sulle pratiche didattiche dei docenti misurate da un osservatore esterno. Questi effetti sono espressi in termini di *effect size* in 4 studi su 8 (Bos et al., 2012; Garet et al., 2008; Gersten, 2010; Sailors & Price, 2010), mentre in uno studio (Grigg, 2012) non abbiamo una misura comparabile; in Cavalluzzo et al. (2012) gli effetti del trattamento sono espressi per alcune misure in forma di *effect size*, mentre per altri in forma di cambiamento percentuale. Al fine della computazione dei risultati non è stato inserito Matsamura et al. (2010), in cui viene riportata la significatività dei risultati, senza riportare coefficienti o *effect size*. Matsamura et al. (2012), invece, riporta gli effetti del trattamento dopo un anno dalla conclusione del trattamento che, per questo motivo, saranno commentati separatamente.

Nel commentare i risultati bisogna tenere in mente la grande eterogeneità dei protocolli di osservazione adottati, che non sempre rilevano le stesse dimensioni. Negli studi citati viene sempre utilizzato uno strumento differente, talvolta sviluppato *ad hoc*. Il concetto di "qualità dell'insegnamento" varia a seconda dello strumento, così come l'enfasi data a specifiche azioni degli insegnanti. Si può tuttavia affermare che in linea generale questi strumenti cercano di misurare la distanza che esiste tra l'insegnamento cosiddetto tradizionale (basato principalmente su lezioni frontali e poco dialogate) con approcci che in maniera efficace coinvolgano gli studenti nella lezione in maniera propositiva.

In 7 casi su 8, la formazione è riuscita a incidere sul modo di insegnare. Ad eccezione di Bos et al. (2012), gli insegnanti paiono aver tratto giovamento dalla formazione in quanto utilizzano meno tempo in classe nella spiegazione frontale, privilegiando l'utilizzo di domande stimolo (Cavalluzzo et al., 2012), espongono i contenuti delle lezioni di scienze in modo da stimolare negli studenti domande e ragionamenti che presuppongano l'utilizzo del metodo scientifico (Grigg, 2012), ottengono un più alto punteggio nell'indice SIOP (Sailors & Price, 2010) o IQA (Matsamura et al., 2010; 2012), e un maggior grado di chiarezza nella fase di spiegazione (Garet et al., 2008) o, in generale, una più alta qualità della discussione insegnante-studenti a lezione (Gersten et al., 2010).

### Atteggiamenti verso l'insegnamento e pratiche didattiche auto-dichiarate

In 7 studi sono riportati effetti sugli atteggiamenti: in 3 casi (Heller et al., 2012; Masters et al., 2010; Greenleaf et al., 2011) sono riportati effetti positivi della formazione su alcune batterie di item riguardanti un vario insieme di atteggiamenti verso l'insegnamento e la frequenza di utilizzo di pratiche didattiche. In 2 casi si assiste a una commistione di effetti nulli ed effetti statisticamente significativi (Argentin et al., 2014; Cavalluzzo et al., 2012), mentre in due casi abbiamo casi di effetti nulli (Bos et al., 2012; Arens et al., 2012).

Nel caso degli atteggiamenti e delle pratiche didattiche autodichiarate ci troviamo di fronte a un'eterogeneità ancora maggiore rispetto alle pratiche didattiche osservate, a cui si affiancano misure self-report sulla percezione di autoefficacia, la soddisfazione lavorativa, gli effetti percepiti della formazione<sup>10</sup> e, più in generale, gli atteggiamenti verso l'insegnamento.

In Arens et al. (2012), non sono stati trovati effetti sulle batterie relative a un vasto insieme di pratiche didattiche, così come in Argentin et al. (2014) e in Bos et al. (2012). Al contrario, in Masters et al. (2010) e Greenleaf et al. (2011) i docenti riportano di utilizzare maggiormente in classe domande stimolo e uno stile di insegnamento meno centrato sulla lezione frontale.

L'effetto dei programmi di formazione sulla percezione di autoefficacia è stimata in Argentin et al. (2014) e Heller et al. (2012). I programmi di formazione hanno un effetto positivo sulle "teacher efficacy beliefs" in Heller et al. (2012), mentre in Argentin et al. (2014) l'effetto complessivo sulla batteria di autoefficacia è nullo, ma si rileva un effetto negativo nell'item relativo alla possibilità di far lavorare gli studenti assieme. Sebbene ispirate dalla teoria della *social cognition* di Bandura (1997), le scale divergono negli item proposti. Nel questionario somministrato ai docenti dell'esperimento di Argentin et al. (2014), è stato stimato l'effetto anche su batterie circa

---

<sup>10</sup> L'effetto dei benefici percepiti del PD è stimato solo quando il programma di cui testare l'efficacia era contrapposto a una versione "base" dello stesso, fornita ai membri del gruppo di controllo (come ad esempio in Cavalluzzo et al., 2012).

la collaborazione dei docenti all'interno della scuola, l'atteggiamento dei docenti rispetto all'insegnamento della matematica e la soddisfazione lavorativa. In queste batterie l'effetto riscontrato è generalmente nullo, a eccezione di alcuni item relativi alla collaborazione interna alla scuola con i colleghi.

#### **4.2.2. Effetti sugli studenti**

Il dispiegarsi di effetti sugli studenti è lo scopo ultimo della realizzazione di interventi di sviluppo professionale per gli insegnanti. Si potrebbe considerare quale *effetto* qualsiasi tipo di ricaduta che un intervento produce sugli studenti degli insegnanti formati: atteggiamento verso lo studio, atteggiamento nei confronti degli insegnanti e dei compagni di classe, risultati scolastici, comportamenti fuori dalla classe per citarne alcuni. Tuttavia nell'illustrare i risultati che emergono dall'insieme degli studi considerati in questa rassegna, sarà presa in considerazione unicamente la *ricaduta della formazione sull'apprendimento degli studenti*. Le ragioni di questa scelta sono piuttosto semplici: innanzitutto perché questo è il focus della nostra ricerca; in seconda istanza perché le misure di *outcome* più affidabili (e comparabili) sono quelle relative ai punteggi ottenuti in appositi test standardizzati.

#### *Le possibili variabili di mediazione*

Prima di presentare i risultati numerici, vale la pena di soffermarci sui possibili meccanismi che rendono possibile la trasformazione di quanto appreso dagli insegnanti –in termini di conoscenze e/o competenze – in miglioramento delle performance degli studenti. Allo stato attuale la ricerca non sembra tracciare un chiaro quadro delle variabili di mediazione. Gli studi sulla formazione degli insegnanti che si occupano di questo particolare aspetto tendono, infatti, a utilizzare per lo più misure a breve termine (Burke & Hutchins, 2007) e di tipo self-report (Blume et al., 2010), aspetti che non consentono di valutare in modo adeguato le ricadute dei programmi in termini di effettivo trasferimento delle competenze apprese. Un altro aspetto critico riguarda il focus per lo più esclusivo sul successivo utilizzo nel contesto lavorativo delle conoscenze e delle competenze apprese (*use measures of transfer*), piuttosto che sull'efficacia di tale utilizzo (*effectiveness measures of transfer*). Tuttavia, tra le variabili di mediazione maggiormente esaminate vengono riportate le abilità generali/trasversali (*open skills*) del docente, dal momento che l'insegnamento è per lo più connesso alla capacità di confrontarsi con la variabilità, piuttosto che all'implementazione di tecniche standardizzate o di pratiche routinarie (Darling-Hammond & Snyder, 2000).

De Rijdt, Stes, Van der Vleuten, e Dochy (2012) hanno proposto un aggiornamento del quadro teorico relativo ai fattori predittivi o di mediazione coinvolti nel trasferimento agli studenti delle competenze apprese dagli insegnanti, a seguito di una rassegna sugli interventi di formazione professionale rivolti a docenti. Gli ulteriori fattori che vengono annoverati sono: l'esperienza del docente (caratteristiche dei formandi), la natura della formazione, la durata del percorso formativo e il clima di apprendimento (caratteristiche della formazione). In particolare, alcuni risultati preliminari della rassegna mostrano che:

- i docenti con più elevata esperienza professionale riescono a trasferire le formazione acquisita nel proprio contesto lavorativo in misura maggiore di quanto facciano i colleghi novizi<sup>11</sup> che mostrano, invece, maggiore capacità di trasferire quanto appreso quando hanno la possibilità di collaborare con docenti più esperti;
- una tipologia di formazione *on the job* (es. comunità di pratiche, gruppi studio, ricerca-azione, coaching o mentoring, studio di caso attraverso la produzione di diari di bordo) risulta più efficace di una tipologia di formazione più tradizionale fruita al di fuori dal contesto scolastico (seminari, workshop, corsi) in quanto i docenti hanno la possibilità di sperimentare direttamente e confrontarsi con problemi e situazioni lavorative autentiche;
- una formazione continua risulta più efficace di interventi formativi spot o che prevedono una singola sessione;
- un clima di apprendimento maggiormente orientato a valorizzare le competenze degli insegnanti e a migliorare le soluzioni già esistenti (*appreciative approach*) risulta associato a outcome migliori rispetto ad un clima centrato sull'individuazione di ciò che non funziona entro l'organizzazione e sull'implementazione di un piano di azione (*gap approach*).

#### **4.2.3 Gli effetti complessivi e per disciplina**

Dei 31 studi inclusi nella rassegna sono 28 quelli che stimano gli effetti della formazione professionale degli insegnanti sull'apprendimento degli studenti. Riportiamo una sintesi dei risultati utilizzando gli *effect size* che sono misure adimensionali (cioè non dipendono dalla metrica del test che è stata utilizzata) e quindi confrontabili: qualora in un singolo studio fossero state utilizzate più misure per stimare l'effetto sulla stessa dimensione – abilità in lingua o in matematica/scienze - o misurate molte dimensioni diverse di una stessa dimensione, prima di calcolare la media generale è stato calcolato l'effetto medio per quel singolo studio (ciò al fine di evitare che gli studi che includevano molte misure potessero pesare maggiormente sui risultati complessivi della rassegna).

---

<sup>11</sup> Sono classificati come novizi gli insegnanti che hanno meno di 5 anni di esperienza professionale.

L'effetto medio sull'apprendimento degli studenti è 0,27. Più di un quarto di deviazione standard che è un ordine di grandezza considerato buono negli studi in ambito educativo (“*educationally significant*”).

Tre studi mostrano un effetto negativo, ma piccolo e statisticamente non significativo, uno mostra un effect size pari a zero, mentre gli altri 24 variano da 0,02 a 0,89;

Quindici studi hanno misurato gli effetti della formazione sull'apprendimento della matematica o delle scienze: il loro *effect size* medio è di 0,23;

Leggermente più alto (0,30) è l'effetto medio dei 19 studi che hanno guardato all'apprendimento della lingua (lettura, scrittura, comprensione del testo o ricchezza del vocabolario). Questo risultato è influenzato dai tre studi focalizzati sull'apprendimento della lingua per studenti non madrelingua o studenti in difficoltà il cui *l'effect size* medio risulta essere pari a 0,61

I quattro percorsi di sviluppo professionale che si proponevano di avere ricadute su lingua e matematica hanno fatto rilevare effetti importanti e statisticamente significativi in entrambe le aree di conoscenza: l'effetto medio è stato di 0,26 con un picco di 0,47 per l'apprendimento della matematica rilevato da Lee. Una spiegazione possibile per il successo in entrambe le discipline è dato dalla – piuttosto nota - difficoltà di comprensione del linguaggio scientifico prima ancora dei contenuti che esso media (soprattutto da parte di studenti deboli): in questo caso uno sviluppo professionale che fonde l'approccio linguistico e disciplinare insieme riesce ad essere molto efficace sulla matematica e le scienze, perché ha rimosso anche alcune barriere linguistiche.

*In sintesi l'effetto sugli studenti è positivo nella gran parte degli studi (anche se in alcuni casi è molto piccolo), ed è leggermente più grande quando la formazione professionale si propone di avere effetti sull'apprendimento della lingua. Quando è negativo, non è statisticamente significativo (ciò indica semplicemente che la formazione professionale ricevuta dagli insegnanti non ha avuto effetti sugli studenti). Gli effetti sugli studenti in difficoltà sono molto più grandi anche se è una indicazione che sembra prematuro generalizzare, perché basata su soli tre studi.*

#### **4.2.4 Effetto per grado di scuola**

Circa metà degli studi considerati (16) riguardano esclusivamente la scuola primaria (dalle scuole per l'infanzia fino alla quinta elementare); due comprendono insegnanti (e studenti) sia della primaria che della secondaria di primo grado (gradi da 6 a 8); 9 solo la scuola secondaria di primo grado, 2 solo la scuola superiore (gradi da 9 a 12) e 1 sia la media che la superiore.

L'effetto medio per gli studi a livello della scuola elementare è di 0,38 deviazioni standard per gli studi su matematica/scienze e 0,36 per quelli di lingua, quindi piuttosto alto e molto omogeneo

per area disciplinare. Gli altri gradi di scuola registrano effetti più bassi. Nella scuola media l'effetto sulla matematica (4 studi) è prossimo allo 0 (0,03), un risultato medio ottenuto combinando effetti (deboli) positivi con negativi; l'effetto medio sulle competenze linguistiche (5 studi) è di 0,20 circa e con risultati coerenti gli uni con gli altri. Solo due studi si focalizzano sulla scuola superiore: gli effetti sono vicini allo zero per la matematica e di circa 0,20 per la lingua.

#### **4.2.5 Effetti per durata del corso**

Per durata del corso intendiamo il numero totale di ore di formazione previsto dai protocolli di formazione professionale previsti dagli studi. All'interno del computo sono calcolate le ore di lezione (frontale e on-line), le attività di *coaching*, seminari di approfondimento estivi e durante l'anno scolastico. Non sono quindi contate le ore di sperimentazione in classe, né le ore di preparazione individuale di un docente.

Il numero medio di ore per corso si attesta su una media di 56, variando tra un minimo di 10 e un massimo di 114. In quattro studi su ventinove il numero di ore non è specificato chiaramente e pertanto questi studi non sono stati inseriti. Se consideriamo che normalmente una giornata di formazione dura l'equivalente di 6 ore lavorative, possiamo distinguere i corsi di formazione in quattro gruppi: corsi che durano l'equivalente di una settimana o meno (analizzati in 7 studi); corsi che durano tra una e due settimane (8 studi); corsi che durano tra due e tre settimane (4 studi); corsi che durano oltre tre settimane (5 studi). Viene di seguito riportato *l'effect size* medio per ogni tipo di durata.

- 1 settimana: 0,29 (variando tra 0,07 e 0,64);
- 2 settimane: 0,26 (variando tra -0,03 e 0,85);
- 3 settimane: 0,18 (variando tra 0,05 e 0,41);
- oltre 3 settimane: 0,2 (variando tra 0 e 0,39).

Come si può notare, la durata in ore di un corso non sembra essere un elemento discriminante per la loro efficacia. Le differenze tra gruppi non seguono un andamento lineare e si può osservare una forte variabilità intra-gruppo.

*L'implicazione che deriva da quest'analisi è che la durata del corso – in sé – non sembra essere un elemento determinante rispetto agli effetti generati sugli studenti.*

#### **4.2.6 Effetti secondo la classificazione di Kennedy**

I programmi sono stati differenziati per contenuto dei corsi in base alle quattro tipologie proposte da Kennedy (1998) – illustrate nel paragrafo 2. Dei 28 percorsi di formazione che riportano effetti sugli studenti, 2 si possono ricondurre alla prima famiglia (didattica generale), 12

alla seconda (didattica disciplinare), 9 alla terza (insegnamento del curriculum) e 5 alla quarta (legata ai processi di apprendimento). L'idea dietro questa classificazione è semplice: cercare di stabilire una connessione tra gli effetti misurati sugli studenti e la teoria del cambiamento sottostante la formazione ricevuta dagli insegnanti, per determinare se e quale tipo di contenuto produca maggiori ricadute sull'apprendimento degli studenti.

I due corsi di formazione della prima famiglia hanno prodotto un *effect size* medio di 0,48; quelli della seconda un *effect size* medio di 0,18; quelli della terza un *effect size* medio di 0,20; infine, quelli della quarta famiglia hanno prodotto un *effect size* medio di 0,43. Le famiglie 1 e 4 (dove esistono *effect size* medi più elevati) sono anche abbastanza omogenee al loro interno: l'*effect size* calcolato in ognuno di questi studi è uguale o superiore a 0,2. Molto più eterogenee sono le altre due famiglie; in particolare, la seconda famiglia è quella in cui si concentra il maggior numero di effetti nulli sugli studenti. Si tratta di risultati abbastanza sorprendenti: i due tipi di corsi maggiormente orientati verso le prove di apprendimento sono anche quelli che registrano minori progressi sugli studenti. Una possibile spiegazione deriva dal semplice fatto che le due famiglie su cui si riscontrano forti effetti positivi hanno numerosità molto ridotte: solamente 7 studi, 2 nella prima e 5 nella quarta famiglia (la quale, infatti, presenta una variabilità discreta).

## Conclusioni

L'evidenza empirica sulla formazione professionale e sulla valutazione della sua efficacia qui presentata, delinea un quadro di notevole incertezza rispetto alle domande che hanno guidato questa rassegna. Nonostante la discreta numerosità di studi (e relativi *outcome*) considerati non è possibile trarre delle indicazioni, neppure di massima, circa gli elementi chiave che caratterizzano una formazione professionale efficace e che possano essere utilizzati per guidare il disegno e l'implementazione di futuri interventi. Ciò è dovuto alla notevole eterogeneità che caratterizza gli interventi valutati. La lettura dei risultati permette tuttavia di individuare alcuni elementi utili per ripensare la ricerca futura: proponiamo di seguito i cinque a nostro avviso più rilevanti.

*1. Ci sono progressi nella ricerca sulla formazione degli insegnanti, ma siamo solo ai primi passi.*

La prima – e, sino ad ora unica - rassegna estensiva sugli effetti della formazione degli insegnanti di Yoon pubblicata nel 2007 aveva individuato solamente 9 studi rigorosi realizzati nei precedenti 25 anni. A pochi anni di distanza è stato possibile, seguendo gli stessi criteri, individuarne altri 24. Ciò significa che c'è stato movimento in questa direzione anche grazie all'impulso del *Department of Education* americano che ha investito ingenti risorse stimolando la

realizzazione di ricerche in grado di fornire evidenza robusta su pratiche concrete efficaci. Un solo studio tra quelli che rispondevano ai criteri per essere inclusi nella rassegna non è americano, ed è stato realizzato in Italia. Può darsi che altri studi siano stati realizzati in Europa, ma non risultano nelle banche dati della produzione scientifica internazionale indicizzata nei motori di ricerca in tema di educazione e formazione utilizzati nel nostro lavoro per il periodo in esame (2000-2012). In ogni caso i 32 studi analizzati sono ancora pochi per trarre conclusioni convincenti, soprattutto in merito ai meccanismi (variabili di mediazione o caratteristiche), che possono determinare il successo di un percorso di formazione professionale per insegnanti.

## *2. Il meccanismo di trasmissione degli effetti (da docenti a studenti) non è ancora chiaro.*

Uno degli obiettivi della rassegna era cercare di delineare un quadro di insieme degli effetti della formazione sui docenti e la relazione tra questi e gli effetti sugli studenti. L'estrema frammentazione degli *outcome* su cui vengono rilevati gli effetti sugli insegnanti, tuttavia, impedisce di tracciare delle linee interpretative. Nel proporre i modelli teorici o concettuali alla base della formazione degli insegnanti, infatti, gli autori degli studi si rifanno a una pluralità di paradigmi e di modelli teorici che raramente possono essere tradotti empiricamente con l'utilizzo di indicatori confrontabili per valutare l'effetto sugli insegnanti. Non vi è nemmeno accordo su quali siano le dimensioni a cui fare riferimento. Esiste un generale accordo sul fatto che gli effetti sugli insegnanti siano da considerarsi "*outcome* intermedi" nel processo che porta un corso di formazione a incidere sugli studenti, ma non è ancora ben chiaro quale sia il percorso che porta dagli uni agli altri o, meglio detto, quali canali vengano attivati grazie a un corso di formazione, e come misurare questi cambiamenti.

Operare questa distinzione è anche fondamentale al fine di comprendere quali effetti siano rilevanti e quali, invece, possiamo considerare secondari: esistono numerosi studi che mostrano un insieme di effetti nulli e positivi, ma senza una chiara mappa concettuale non è possibile operare un chiaro discrimine tra gli effetti sugli insegnanti su cui occorre concentrarsi e quelli di carattere esplorativo o accessorio.

## *3. La formazione degli insegnanti può essere efficace (e anche molto). Ci sono caratteristiche che danno garanzie di maggior efficacia?*

La gran parte degli studi analizzati mostra un effetto positivo sugli studenti (in 24 dei 29 studi che hanno rilevato gli effetti sugli studenti), anche se spesso si tratta di un effetto molto piccolo o prossimo allo zero. Se da un lato questo è un risultato confortante, in quanto conferma che migliorare conoscenze e competenze degli insegnanti si traduce spesso in miglioramento delle performance degli studenti, dall'altro i risultati confermano quanto risulti difficoltoso evincere quali

possano essere i fattori di successo dei programmi efficaci rispetto a quelli poco o per nulla efficaci. Sono pochi i casi di interventi che si soffermano sull'individuazione delle caratteristiche che hanno potuto contribuire al successo del programma. Anche quando questo accade, tuttavia, la discussione rimane confinata all'interno del singolo caso sotto esame, e non viene discussa un'eventuale applicabilità ad altri casi, in cui corsi interventi con aventi le stesse caratteristiche non hanno prodotto effetti. Caso emblematico è lo studio di Heller et al. (2012b) che utilizza un esperimento con disegno fattoriale che vede ben tre gruppi di trattamento confrontati con un gruppo di controllo. In tutti e tre i gruppi di trattamento lo stesso argomento (come insegnare elementi di base dei circuiti elettronici) viene affrontato in modi radicalmente diversi. Purtroppo, gli effetti dei tre trattamenti non differiscono tra di loro (se non in maniera marginale), il che impedisce agli autori di trarre delle conclusioni circa la preferibilità di un'opzione formativa rispetto alle altre.

L'evidenza prodotta dalla nostra analisi suggerisce comunque tre risultati chiave da cui partire comunque che:

- negli studi considerati la formazione aveva una durata (o un'intensità) media o alta. All'interno di questi corsi quindi, la durata non è discriminante per l'efficacia dei programmi. La durata (l'intensità) non ha fatto differenza, confermando quanto già aveva rilevato una ricerca di Kennedy (1998). Tuttavia occorre specificare che i percorsi molto brevi erano stati esclusi a priori (seminari, incontri, workshop e interventi formativi di una sola giornata o di durata complessiva inferiore a 10 ore), in quanto tutta la ricerca precedente li aveva già con certezza etichettati come inefficaci;
- gli interventi formativi realizzati nella scuola primaria hanno un effetto medio molto più alto rispetto a quelli realizzati nei gradi di scuola successivi (soprattutto per quanto riguarda le competenze matematiche), confermando i risultati di un'ampia letteratura sui processi educativi che mostra come sia più efficace, attraverso i più svariati canali, agire quando gli studenti sono ancora in tenera età. Senza voler escludere gli insegnanti e gli studenti delle scuole superiori, se si vuole davvero incidere sulle competenze degli studenti occorre agire con decisione nelle scuole primarie;
- negli studi che rilevano *outcome* di vario tipo sia su docenti sia su studenti si osserva una forte associazione tra effetti sugli insegnanti e effetti sugli studenti. Quando i primi sono positivi allora si registrano effetti anche sugli studenti. Viceversa, gli studi che mostrano effetti piccoli, negativi o nulli sui docenti non producono effetti nemmeno sugli studenti. Questo risultato ricorda quanto sia importante porre attenzione alla *teoria del cambiamento* soggiacente al programma valutato e pianificare di conseguenza le rilevazioni opportune sugli insegnanti, se si

vuole capire di più su come l'intervento funzioni. Queste, infatti, non possono più essere considerate accessorie.

#### *4. Ancora non è possibile avanzare ipotesi su quali siano le variabili di mediazione più rilevanti.*

Poco chiaro rimane il legame tra quali siano i cambiamenti nel modo di insegnare degli insegnanti associati a miglioramenti nell'apprendimento degli studenti. Questa domanda è di difficile risposta, in quanto non esiste un modo condiviso di operativizzare i costrutti sottostanti alle dimensioni relative al lavoro in classe degli insegnanti, e le stesse griglie di codifica non si rifanno sempre alle stesse dimensioni. Riuscire a stabilire un legame in questo senso aumenterebbe molto la comprensione di cosa funziona nella formazione professionale, anche se, come abbiamo visto, le dimensioni lungo le quali gli interventi differiscono sono molte e l'evidenza disponibile non permette di operare un chiaro discrimine tra pratiche efficaci e pratiche inefficaci. Anche da questo punto di vista, è possibile che un'attenzione maggiore ai modi in cui gli interventi sono stati realizzati nel concreto, e all'analisi del processo di trasferimento di quanto appreso nella formazione porti a chiarire maggiormente i legami esistenti tra formazione, cambiamenti nelle pratiche e risultati degli studenti.

#### *5. Possono essere le caratteristiche stesse dello studio a compromettere la possibilità di rilevare gli effetti.*

Gli autori degli studi che hanno rilevato effetti negativi o nulli, hanno ipotizzato alcune spiegazioni di questi risultati che vanno al di là del contenuto della formazione, ma riguardano le modalità stesse di realizzazione del programma e dello studio ad esso associato. La più convincente di queste ipotesi ha a che fare con la tempistica delle rilevazioni degli effetti, forse uno dei principali limiti in cui incorrono gli studi valutativi qui presi in rassegna. Gli interventi valutati hanno spesso una durata di circa un anno scolastico, nel corso del quale gli insegnanti devono familiarizzare, interiorizzare ed imparare ad utilizzare i contenuti della formazione. Di norma gli effetti sugli studenti vengono rilevati a fine anno scolastico, forse troppo presto perché gli insegnanti possano padroneggiare al meglio i nuovi contenuti o metodi proposti e, di conseguenza, trasferirne il potenziale positivo sugli studenti.

Questa considerazione è particolarmente calzante quando si tratta di rilevare degli effetti sugli insegnanti. Nel caso delle applicazioni di protocolli standardizzati di osservazione in classe, normalmente si osserva l'effetto durante il periodo in cui i docenti sono esposti alla formazione. Nella maggior parte dei casi è difficile capire se le pratiche didattiche vengano incorporate o meno all'interno del bagaglio professionale dell'insegnante nel medio-lungo periodo. La distinzione tra

effetti di breve e medio-lungo periodo è fondamentale: un effetto di breve periodo sui docenti non solo non permette di parlare di “sviluppo professionale” in senso proprio (in quanto i benefici potrebbero rivelarsi effimeri e non portare un valore aggiunto al docente e alla scuola negli anni seguenti), ma potrebbe anche non trattarsi propriamente di un “vero” effetto della formazione sul personale docente. Detto altrimenti, i cambiamenti nei comportamenti didattici rilevati dalle analisi potrebbero essere, semplicemente, il riflesso dell’implementazione in classe del protocollo di sperimentazione in cui i docenti dei vari gruppi di sperimentazione erano inseriti. Non si tratta tuttavia di un semplice problema metodologico: le pochissime ricerche che hanno stimato effetti a due anni di distanza mostrano come siano declinanti nel tempo. Questo risultato, se confermato da altre ricerche, potrebbe indurre a rivedere i protocolli di intervento attualmente previsti per molti interventi formativi affinché all’insegnante possa essere garantita continuità negli anni. Seguendo quest’ottica, più vicina a un modello di *life-long-learning* dei modelli esistenti, nuovi metodi, tecniche di insegnamento e/o conoscenze potrebbero sedimentarsi e diventare compiutamente parte del bagaglio professionale degli insegnanti, per beneficio loro e delle scuole nelle quali insegnano.

## Riferimenti Bibliografici

I riferimenti indicati con l'asterisco (\*) indicano gli studi inclusi nell'analisi degli effetti

Angrist, J., Lavy, V. (2001). The Effect of Teachers' Training on Student Achievements, *Journal of Labor Economics*, 19 (2), pp. 343-369.

Antoniou, P. & Kyriakides, L. (2013). A Dynamic Integrated Approach to teacher professional development: Impact and sustainability of the effects on improving teacher behaviour and student outcomes. *Teaching and Teacher Education*, 29, 1–12. doi: 10.1016/j.tate.2012.08.001

\*Arens, S. A., Stoker, G., Barker, J., Shebby, S., Wang, X., Cicchinelli, L. F., & Williams, J. M. (2012). *Effects of Curriculum and Teacher Professional Development on the Language Proficiency of Elementary English Language Learner Students in the Central Region*. Denver: COMid-continent Research for Education and Learning. ED530839.

\*Argentin, G., Pennisi, A., Vidoni, D., Abbiati, G., & Caputo, A. (2014). Trying to Raise (Low) Math Achievement and to Promote (Rigorous) Policy Evaluation in Italy Evidence From a Large-Scale Randomized Trial. *Evaluation Review*, 0193841X14529125. doi:10.1177/0193841X14529125

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Berk, R. (2005). Randomized experiments as the bronze standard, *Journal of Experimental Criminology*, 4 (1), 417—433.

Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of Management*, 36(4), 1065-1105.

\*Booth Olson, C., Kim, J. S., Scarcella, R., Kramer, J., Pearson, M., Dyk, D. A. van, Land, R. E. (2012). Enhancing the Interpretive Reading and Analytical Writing of Mainstreamed English Learners in Secondary School Results From a Randomized Field Trial Using a Cognitive Strategies Approach. *American Educational Research Journal*, 49(2), 323–355. doi:10.3102/0002831212439434

\*Bos, J. M., Sanchez, R. C., Tseng, F., Rayyes, N., Ortiz, L., & Sinicrope, C. (2012). Evaluation of Quality Teaching for English Learners (QTEL) professional development (NCEE 2012-4005). *Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education*.

Burke, L. A., & Hutchins, H. M. (2007). Training transfer: An integrative literature review. *Human Resource Development Review*, 6(3), 263-296.

Calderhead, J. (1989). Reflective teaching and teacher education. *Teaching and Teacher Education*, 5, 43-51.

\*Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C.-P., & Loef, M. (1989). Using Knowledge of Children's Mathematics Thinking in Classroom Teaching: An Experimental Study. *American Educational Research Journal*, 26(4), 499–531. doi:10.3102/00028312026004499

\*Cavalluzzo, L., Lowther, D., Mokher, C., and Fan, X. (2012). *Effects of the Kentucky Virtual Schools' hybrid program for algebra I on grade 9 student math achievement*. (NCEE 2012-4020). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Clift, R. T., Houston, W. R., & Pugach, M. C. (Eds.) (1990). *Encouraging reflective practice: An analysis of issues and programs*. New York: Teachers College.

Clotfelter, C., Ladd, H., Vigdor, J. (2007). Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects, *CALDER working paper* 11/2007.

\*Cole, D. C. (1992). The effects of a one-year staff development program on the achievement of test scores of fourth-grade students. *Dissertation Abstracts International*, 53(06), 1792A. (UMI No. 9232258)

Commissione Europea (2004). *Common European Principles for Teacher Competences and Qualifications*, Bruxelles, Commissione Europea, disponibile all'indirizzo [http://ec.europa.eu/education/policies/2010/doc/principles\\_en.pdf](http://ec.europa.eu/education/policies/2010/doc/principles_en.pdf).

Correnti, R. (2007). An Empirical Investigation of Professional Development Effects on Literacy Instruction using Daily Logs, Educational Evaluation and Policy Analysis, 29 (4), pp. 262–295.

Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and teacher education*, 16(5-6), 523-545.

De Rijdt, C., Stes, A., Van der Vleuten, C., & Dochy, F. (2013). Influencing variables and moderators of transfer of learning to the workplace within the area of staff development in higher education: Research review. *Educational Research Review*, 8, 48-74. doi:10.1016/j.edurev.2012.05.007

Desimone, L. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures, *Educational Researcher*, 38 (3), pp. 181-199.

Desimone, L., Porter, A., Garet, M., Yoon, K.S., Birman, B. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study, *Educational Evaluation and Policy Analysis*, 24 (2), pp. 81-112.

\*Duffy, G. G., Roehler, L. R., Meloth, M. S., Vavrus, L. G., Book, C., Putnam, J., & Wesselman, R. (1986). The relationship between explicit verbal explanations during reading skill instruction and student awareness and achievement: A study of reading teacher effects. *Reading Research Quarterly*, 21(3), 237–252.

Elliot, J. (2002). *Action research as the basis of a new professionalism for teachers in an age of globalisation*. Paper presented at a conference to celebrate the centenary year of Beijing Normal University, Beijing, China.

Eurydice (2000). *European trends in anticipation of teacher training needs – Summary of answers from the Eurydice network. Anticipatory project to investigate teachers' initial and continuing training needs* (OPEPRO), report 11, Finland, National Board of Education.

Eurydice (2006). *Quality assurance of teacher education in Europe*, Bruxelles, Eurydice.

Ganser, T. (2007). An ambitious vision of professional development for teachers, *NASSP Bulletin*, 84 (618), pp. 6-12.

Garet, M., Porter, A., Desimone, L. Birman, B., Yoon, K. (2001). What makes professional development effective? Analysis of a national sample of teachers, *American Education Research Journal*, 38 (4), pp. 915-945.

\*Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., Uekawa, K., et al. (2008). *The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement*. NCEE 2008-4030.

\*Garet, M. S., Wayne, A. J., Stancavage, F., Taylor, J., Eaton, M., Walters, K., Song, M., et al. (2011). *Middle School Mathematics Professional Development Impact Study: Findings after the Second Year of Implementation*. Executive Summary. NCEE 2011-4025. National Center for

Education Evaluation and Regional Assistance. Retrieved from <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED519923>

\*Gersten, R., Dimino, J., Jayanthi, M., Kim, J. S., & Santoro, L. E. (2010). *Teacher Study Group: Impact of the Professional Development Model on Reading Instruction and Student Outcomes in First Grade Classrooms*. *American Educational Research Journal*, 47(3), 694-739.

Gilberts, G., & Lignugaris-Kraft, B. (1997). Classroom management and instruction competencies for preparing elementary and special education teachers. *Teaching and Teacher Education*, 13(6), 597-610.

Glatthorn, A. (1995). *Teacher development*, in Anderson, L. (a cura di), *International encyclopedia of teaching and teacher education*, Londra, Pergamon Press.

\*Glazerman, S., Dolfen, S., Bleeker, M., Johnson, A., Isenberg, E., Lugo-Gil, J., ... Ali, M. (2008). Impacts of Comprehensive Teacher Induction: Results from the First Year of a Randomized Controlled Study. NCEE 2009-4034. *National Center for Education Evaluation and Regional Assistance*. Retrieved from <http://eric.ed.gov/?id=ED503061>

\*Greenleaf, C. L., Litman, C., Hanson, T. L., Rosen, R., Boscardin, C. K., Herman, J., & Schneider, S. A. (2011). Integrating Literacy and Science in Biology: Teaching and Learning Impacts of Reading Apprenticeship Professional Development. *American Educational Research Journal*, 48(3), 647-717.

\*Grigg, J., Kelly, K. A., Gamoran, A., & Borman, G. D. (2012). Effects of two scientific inquiry professional development interventions on teaching practice. *Educational Evaluation and Policy Analysis*, 0162373712461851.

Guskey, T. (2009). Closing the Knowledge Gap on Effective Professional Development, *Educational Horizons*, 87 (4), pp. 224-233.

Jacob, B.A., Lefgren, L. (2004). *The Impact of Teacher Training on Student Achievement: Quasi-Experimental Evidence from School Reform Efforts in Chicago*, *The Journal of Human Resources*, 39 (1), pp. 50-79.

Kennedy, M. (1998). Form and substance of in-service teacher education, Madison, WI, National Institute for Science Education, University of Wisconsin–Madison.

\*Heller, J. I., Daehler, K. R., Wong, N., Shinohara, M., & Miratrix, L. W. (2012). Differential effects of three professional development models on teacher knowledge and student achievement in elementary science. *Journal of Research in Science Teaching*, 49(3).

\*Kim, J. S., Olson, C. B., Scarcella, R., Kramer, J., Pearson, M., van Dyk, D., Collins, P., et al. (2011). A Randomized Experiment of a Cognitive Strategies Approach to Text-Based Analytical Writing for Mainstreamed Latino English Language Learners in Grades 6 to 12. *Journal of Research on Educational Effectiveness*, 4, 231-263.

\*Lee, O., Maerten-Rivera, J., Penfield, R. D., LeRoy, K., & Secada, W. G. (2008). Science achievement of English language learners in urban elementary schools: Results of a first year professional development intervention. *Journal of Research in Science Teaching*, 45(1)

\*Marek, E. A., & Methven, S. B. (1991). Effects of the learning cycle upon student and classroom teacher performance. *Journal of Research on Science in Teaching*, 28(1), 41–53.

Martini, A., Sisti, M. (2009), *Valutare il successo delle politiche pubbliche*, Bologna, Il Mulino.

\*Masters, J., de Kramer, R. M., O'Dwyer, L. M., Dash, S., & Russell, M. (2010). The Effects of Online Professional Development on Fourth Grade English Language Arts Teachers' Knowledge and Instructional Practices. *Journal of Educational Computing Research*, 43(3).

- \*Matsumura, L. C., Garnier, H. E., Correnti, R., Junker, B., & Bickel, D. D. (2010). Investigating the Effectiveness of a Comprehensive Literacy Coaching Program in Schools with High Teacher Mobility. *Elementary School Journal*, 111(1), 35-62.
- \*Matsumura, L. C., Garnier, H. E., & Spybrook, J. (2012). The effect of content-focused coaching on the quality of classroom text discussions. *Journal of Teacher Education*, 63(3), 214-228.
- \*McCutchen, D., Abbott, R. D., Green, L. B., Beretvas, S. N., Cox, S., Potter, N. S., Quiroga, T., & Gray, A. L. (2002). Beginning literacy: Links among teacher knowledge, teacher practice, and student learning. *Journal of Learning Disabilities*, 35(1), 69-86.
- \*McGill-Franzen, A., Allington, R. L., Yokoi, L., & Brooks, G. (1999). Putting books in the classroom seems necessary but not sufficient. *Journal of Reading Research*, 93(2), 67-74.
- Miller, E. (1998). The old model of staff development survives in a world where everything else has changed. *Harvard Education Letter*, 11, 2-4.
- Moscatti, R. (2000). *La formazione iniziale e l'aggiornamento*, in Cavalli, A. (a cura di) (2000). *Gli insegnanti nella scuola che cambia. Seconda indagine IARD sulle condizioni di vita e di lavoro nella scuola italiana*, Bologna, Il Mulino.
- OECD, (2005). *Teachers matter: Attracting, Developing and Retaining Effective Teachers*, Paris, OECD.
- OECD (2009), *Creating effective teaching and learning environments - first results from Talis*, Parigi, OECD.
- Phillips, K., Desimone, L., Smith, M. (2011). Teacher Participation in Content-focused Professional Development & The Role of State Policy, *Teachers College Record*, 113, (11), pp. 2586-2621.
- \*Pianta, R. C., Mashburn, A. J., Downer, J. T., Hamre, B. K., & Justice, L. (2008). Effects of web-mediated professional development resources on teacher-child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly*, 23(4), 431-451.
- \*Randel, B., Beesley, A. D., Apthorp, H., Clark, T. F., Wang, X., Cicchinelli, L. F., & Williams, J. M. (2011). *Classroom Assessment for Student Learning: The impact on elementary school mathematics in the Central Region* (NCEE 2011-4005). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- \*Sailors, M., & Price, L. R. (2010). Professional Development that Supports the Teaching of Cognitive Reading Strategy Instruction. *Elementary School Journal*, 110(3), 301-322.
- \*Santagata, R., Kersting, N., Givvin, K. B., & Stigler, J. W. (2010). Problem implementation as a lever for change: An experimental study of the effects of a professional development program on students' mathematics learning. *Journal of Research on Educational Effectiveness*, 4(1), 1-24.
- \*Saxe, G. B., Gearhart, M., & Nasir, N. S. (2001). Enhancing students' understanding of mathematics: A study of three contrasting approaches to professional support. *Journal of Mathematics Teacher Education*, 4, 55-79.
- Scheerens, J. (2000). *Improving School Effectiveness. Fundamentals of Educational Planning*, Paris, Unesco.
- \*Sloan, H. A. (1993). Direct instruction in fourth and fifth grade classrooms. *Dissertation Abstracts International*, 54(08), 2837A. (UMI No. 9334424)
- Sprinthall, N., Reiman A. & Thies-Sprinthall L. (1996). Teacher Professional Development. In J. Sikula (Ed.), *Handbook of Research on Teacher Education* (pp. 666-703). New York: Macmillan.

\*Tienken, C. H. (2003). The effect of staff development in the use of scoring rubrics and reflective questioning strategies on fourth-grade students' narrative writing performance. *Dissertation Abstracts International*, 64(02), 388A. (UMI No. 3081032)

Thomson, P. (1991). *Competency-based training: Some development and assessment issues for policy makers*. Leabrook, Australia: TAFE National Centre for Research and Development.

\*Vaughn, S., Klingner, J. K., Swanson, E. A., Boardman, A. G., Roberts, G., Mohammed, S. S., & Stillman-Spisak, S. J. (2011). Efficacy of Collaborative Strategic Reading With Middle School Students\*. *American Educational Research Journal*, 48(4), 938 -964.

Villegas-Reimers, E. (2003). *Teacher Professional Development: an international review of literature*. Paris: UNESCO/International Institute for Educational Planning.

Wayne, A., Youngs, P. (2003). *Teacher Characteristics and Student Achievement Gains: A Review*, *Review of Educational Research*, 73 (1), pp. 89-122.

Whitty, G., & Willmott, E. (1991). Competence-based teacher education: Approaches and issues. *Cambridge Journal of Education*, 21(3), 309–319.

## APPENDICE A

### Classificazione dei documenti

---

## Classificazione documenti

## APPENDICE A

<b>Autori</b>	<b>Anno pubblicazione</b>	<b>Tipo di documento</b>	<b>Tipologia di campione</b>	<b>Grado di Scuola</b>	<b>Focus di riferimento</b>	<b>Paese di realizzazione</b>	<b>Anni di realizzazione</b>	<b>Tipo di esperienza</b>	<b>Evidenza prodotta</b>
Allen, J., Pianta, R., Gregory, A., Mikami, A., & Lun, J.	2011	Articolo	INS, SCU, STU	9-13	Relazione studenti-insegnanti	Stati Uniti		Studio dimostrativo	Disegno sperimentale
Angrist, J. D., & Lavy V.	2001	Articolo	STU	1-5	Matematica, Lingua nazionale	Israele	1995-1996	Studio dimostrativo	Evidenza controfattuale non sperimentale
Antoniou, P., Kyriakides, L.	2011	Articolo	INS, STU	1-8	Matematica	Cipro	2008-2009	Studio dimostrativo	Altra evidenza quantitativa
Archibald, S., Coggshall, J. G., Croft, A., Goe, L.	2011	Rapporto di ricerca				Stati Uniti	2011		Approfondimento teorico
Arens, S. A., Stoker, G., Barker, J., Shebby, S., Wang, X., Cicchinelli, L. F., & Williams, J. M.	2012	Rapporto di ricerca	INS, SCU, STU	1-5	Lingua nazionale	Stati Uniti	2008-2010	Studio dimostrativo	Disegno sperimentale
Argentin 2014	2014	Articolo	INS, STU	6-8	Matematica	Italia	2009-2012	Studio dimostrativo	Disegno sperimentale
Armour, K. M., & Makopoulou, K.	2012	Articolo	INS	1-13	Altre discipline scientifiche	Inghilterra	2004-2006	Politica a scala	Evidenza qualitativa
Avalos, B.	2011	Articolo	INS						Review/Meta-analisi
Bishop, R., Berryman, M., Wearmouth, J., Peter, M., Clapham, S.	2012	Articolo	STU, INS			Nuova Zelanda	2003-2010	Studio dimostrativo	Altra evidenza quantitativa
Blank, R. K., de las Alas, N.	2009	Rapporto di ricerca	INS, SCU, STU	1-13	Matematica	Stati Uniti	2006-2009		Review/Meta-analisi
Boath-Olson 2012	2012	Articolo	INS, STU	9-13	Lingua nazionale	Stati Uniti		Studio dimostrativo	Disegno sperimentale
Bos, J. M., Sanchez, R. C., Tseng, F., Rayyes, N., Ortiz, L., Sinicrope, C.	2012	Rapporto di ricerca	INS, SCU, STU	6-8	Lingua nazionale	Stati Uniti	2007-2010	Studio dimostrativo	Disegno sperimentale
Boyle, B., Lamprianou, I., & Boyle, T.	2005	Articolo	INS	1-8	Matematica, altre discipline scientifiche	Inghilterra	2002-2003	Politica a scala	Altra evidenza quantitativa

Tipologia di campione: Studenti (STU), Insegnanti (INS), Scuole (SCU)

**Classificazione documenti**

**APPENDICE A**

<b>Autori</b>	<b>Anno pubblicazione</b>	<b>Tipo di documento</b>	<b>Tipologia di campione</b>	<b>Grado di Scuola</b>	<b>Focus di riferimento</b>	<b>Paese di realizzazione</b>	<b>Anni di realizzazione</b>	<b>Tipo di esperienza</b>	<b>Evidenza prodotta</b>
Campbell, C.	2009	Articolo	INS, SCU	Università	Lingua nazionale	Stati Uniti		Politica a scala	Altra evidenza quantitativa
Cantrell, S. C., & Hughes, H. K.	2008	Articolo	INS	6-13	Lingua nazionale	Stati Uniti		Studio dimostrativo	Altra evidenza quantitativa
Capps, D. K., Crawford, B. A., Conostas, M. A.	2012	Articolo	INS		Altre discipline scientifiche				Review/Meta-analisi
Carpenter 1989	1989	Articolo	INS, STU	1-5	Matematica	Stati Uniti		Studio dimostrativo	Disegno sperimentale
Cavalluzzo, L., Lowther, D., Mokher, C., and Fan, X.	2012	Rapporto di ricerca	INS, SCU, STU	9-13	Matematica	Stati Uniti	2007-2009	Studio dimostrativo	Disegno sperimentale
Cheung, A., & Slavin, R.	2012	Rassegna di esperienze	STU	1-5	Lingua nazionale	Stati Uniti			Review/Meta-analisi
Clough, M. P., Berg, C. A., & Olson, J. K.	2009	Articolo							Approfondimento teorico
Cole 1992	1992	Articolo	INS, STU	1-5	Matematica, Lingua nazionale	Stati Uniti	1989-1990	Studio dimostrativo	Disegno sperimentale
Comprehensive School Reform Quality (CSRQ) Center	2006	Rassegna di esperienze	INS, SCU, STU	6-13		Stati Uniti			Review/Meta-analisi
Comprehensive School Reform Quality (CSRQ) Center	2006	Rassegna di esperienze	INS, SCU, STU	1-5		Stati Uniti			Review/Meta-analisi
Correnti, R.	2007	Articolo	INS	1-5	Pratiche didattiche	Stati Uniti	2001-2004	Studio dimostrativo	Evidenza controfattuale non sperimentale
Dealoglu, A.	2004	Articolo	INS	1-5	Altre discipline scientifiche	Turchia	2001-2002	Studio dimostrativo	Altra evidenza quantitativa
DeLaPaz, S., Malkus, N., Monte-Sano, C., Montanaro, E.	2011	Articolo	INS, STU	1-8	Altre discipline scientifiche	Stati Uniti		Studio dimostrativo	Altra evidenza quantitativa

Tipologia di campione: Studenti (STU), Insegnanti (INS), Scuole (SCU)

**Classificazione documenti**

**APPENDICE A**

<b>Autori</b>	<b>Anno pubblicazione</b>	<b>Tipo di documento</b>	<b>Tipologia di campione</b>	<b>Grado di Scuola</b>	<b>Focus di riferimento</b>	<b>Paese di realizzazione</b>	<b>Anni di realizzazione</b>	<b>Tipo di esperienza</b>	<b>Evidenza prodotta</b>
Desimone, L., Porter, A. C., Birman, B. F., Garet, M. S., & Yoon, K. S.	2002	Articolo	SCU	1-13		Stati Uniti		Politica a scala	Altra evidenza quantitativa
Desimone, L.M., Porter, A.C., Garet, M.S., Yoon, K.S. & Birman, B.F.	2002	Articolo	INS	1-13	Pratiche didattiche	Stati Uniti	1996-1999	Studio dimostrativo	Altra evidenza quantitativa
Dingle, M. P., Brownell, M. T., Leko, M. M., Boardman, A. G., & Haager, D.	2011	Articolo	INS	1-5	Lingua nazionale	Stati Uniti		Studio dimostrativo	Evidenza qualitativa
Doğan, N, Çakiroğlu, J., Çavuş, S., Bilican, K., & Arslan, O.	2011	Articolo	INS	1-5	Altre discipline scientifiche	Turchia		Studio dimostrativo	Altra evidenza quantitativa
Dooly, M.	2009	Articolo	INS		ICT	Spagna	2004-2006	Politica a scala	Evidenza qualitativa
Doppelt, Y., Schunn, D. C., Silk, E. M., Mehalik, M. M., Reynolds B. & Ward, E.	2009	Articolo	INS, STU		Altre discipline scientifiche	Stati Uniti	2004-2006	Studio dimostrativo	Altra evidenza quantitativa
Duffy 1986	1986	Articolo	INS, STU	1-5	Lingua nazionale	Stati Uniti		Studio dimostrativo	Disegno sperimentale
Eilks, I., Markic, S.	2011	Articolo	INS		Altre discipline scientifiche	Germania			Evidenza qualitativa
Enochsson, A. & Rizza, C.	2009	Working paper	INS					Studio dimostrativo	Review/Meta-analisi
Fielding, A. & Yang, M.	2005	Articolo	SCU,STU			Inghilterra	1993	Studio dimostrativo	Altra evidenza quantitativa
Franke, M.L., Carpenter, T.P., Levi, L. & Fennema, E.	2001	Articolo	INS			Stati Uniti	1990-1993	Studio dimostrativo	Evidenza qualitativa
Fugate, A.G, Patricia, A.J., & Preuhs R.R.	2001	Articolo	SCU	Università		Stati Uniti	2000	Studio dimostrativo	Approfondimento teorico

Tipologia di campione: Studenti (STU), Insegnanti (INS), Scuole (SCU)

**Classificazione documenti**

**APPENDICE A**

<b>Autori</b>	<b>Anno pubblicazione</b>	<b>Tipo di documento</b>	<b>Tipologia di campione</b>	<b>Grado di Scuola</b>	<b>Focus di riferimento</b>	<b>Paese di realizzazione</b>	<b>Anni di realizzazione</b>	<b>Tipo di esperienza</b>	<b>Evidenza prodotta</b>
Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., Uekawa, K., Falk, A., Bloom, H., Doolittle, F., Zhu, P., & Sztenjnberg, L.	2008	Rapporto di ricerca	INS, SCU, STU	1-5	Lingua nazionale	Stati Uniti	2005-2006	Studio dimostrativo	Disegno sperimentale
Garet, M., Wayne, A., Stancavage, F., Taylor, J., Eaton, M., Walters, K., Song, M., Brown, S., Hurlburt, S., Zhu, P., Sepanik, S., & Doolittle, F.	2011	Rapporto di ricerca	INS, SCU, STU	9-13	Matematica	Stati Uniti	2007-2009	Studio dimostrativo	Disegno sperimentale
Garet, M., Wayne, A., Stancavage, F., Taylor, J., Walters, K., Song, M., Brown, S., Hurlburt, S., Zhu, P., Sepanik, S. & Doolittle, F.	2010	Rapporto di ricerca	INS, SCU, STU	9-13	Matematica	Stati Uniti	2007-2008	Studio dimostrativo	Disegno sperimentale
Garet, M.S., Porter, A.C., Desimone, L., Birman B.F. & Yoon K.S.	2011	Articolo	INS	1-13	Matematica	Stati Uniti	1998	Studio dimostrativo	Altra evidenza quantitativa
Gersten, R., Dimino, J., Jayanthi, M., Kim, J. S., Santoro, L. E.	2010	Articolo	INS, SCU, STU	1-5	Lingua nazionale	Stati Uniti		Studio dimostrativo	Altra evidenza quantitativa
Glazerman, S., Isenberg, E., Dolfin, S., Bleeker, M., Johnson, M., Grider, A., & Jacobus, M.	2010	Rapporto di ricerca	INS, SCU, STU	1-5			2004-2007	Studio dimostrativo	Disegno sperimentale
Greenleaf 2011	2011	Articolo	INS, STU	9-13	Scienze	Stati Uniti	2005-2007	Studio dimostrativo	Disegno sperimentale
Grigg 2012	2012	Articolo	INS, STU	1-5	Scienze	Stati Uniti	2006-2009	Studio dimostrativo	Disegno sperimentale
Hall, T.D. & Scott, R.	2007	Articolo	INS		Altre discipline (Storia)	Stati Uniti	2004	Studio dimostrativo	Evidenza qualitativa

Tipologia di campione: Studenti (STU), Insegnanti (INS), Scuole (SCU)

**Classificazione documenti**

**APPENDICE A**

<b>Autori</b>	<b>Anno pubblicazione</b>	<b>Tipo di documento</b>	<b>Tipologia di campione</b>	<b>Grado di Scuola</b>	<b>Focus di riferimento</b>	<b>Paese di realizzazione</b>	<b>Anni di realizzazione</b>	<b>Tipo di esperienza</b>	<b>Evidenza prodotta</b>
Ham, V.	2010	Articolo	INS		Altre discipline scientifiche	Nuova Zelanda			Review/Meta-analisi
Harris, K. R., Lane, K. L., Graham, S., Driscoll, S. A., Sandmel, K., Brindle, M., Schatschneider, C.	2012	Articolo	INS, SCU, STU	1-5	Lingua nazionale	Stati Uniti		Studio dimostrativo	Altra evidenza quantitativa
Heller, J. I., Daehler, K. R., Wong, N., Shinohara, M., & Miratrix, L. V.	2012	Articolo	STU, INS	1-5	Altre discipline scientifiche, Lingua nazionale	Stati Uniti	2007-2008	Studio dimostrativo	Disegno sperimentale
Heller, J.I.	2012	Rapporto di ricerca	INS, SCU, STU	6-8	Altre discipline scientifiche	Stati Uniti	2009-2010	Studio dimostrativo	Disegno sperimentale
Hoque, K. E., Alam, G. M., Abdullah, Abdul G. K.	2011	Articolo	INS	9-13		Bangladesh		Studio dimostrativo	Altra evidenza quantitativa
Jacob, A.B. & Lefgren, L.	2008	Articolo	INS	1-5	Matematica, Lingua nazionale	Stati Uniti	2003	Studio dimostrativo	Altra evidenza quantitativa
Jacob, B. A., & Lefgren, L.	2004	Articolo	STU, SCU	1-5	Matematica, Lingua nazionale	Stati Uniti	1997-1999	Politica a scala	Evidenza controfattuale non sperimentale
Jaquith, A., Mindich, D., Wei, R. C., Darling-Hammond, L.	2010	Rapporto di ricerca	INS			Stati Uniti			Review/Meta-analisi
Jensen, B.	2010	Capitolo su volume	INS						Approfondimento teorico
Jiang, B., & DeVillar, R. A.	2005	Articolo	INS, STU	6-12 (Scuola dell'infanzia)	Altre discipline scientifiche	Messico		Studio dimostrativo	Altra evidenza quantitativa
Jimoyiannis, A.	2010	Articolo	INS	9-13	Altre discipline scientifiche	Grecia		Politica a scala	Evidenza qualitativa
Johnson C. C., & Fargo, J. D.	2010	Articolo	INS, SCU, STU	6-8	Altre discipline scientifiche	Stati Uniti	2005-2007	Studio dimostrativo	Altra evidenza quantitativa

Tipologia di campione: Studenti (STU), Insegnanti (INS), Scuole (SCU)

**Classificazione documenti**

**APPENDICE A**

<b>Autori</b>	<b>Anno pubblicazione</b>	<b>Tipo di documento</b>	<b>Tipologia di campione</b>	<b>Grado di Scuola</b>	<b>Focus di riferimento</b>	<b>Paese di realizzazione</b>	<b>Anni di realizzazione</b>	<b>Tipo di esperienza</b>	<b>Evidenza prodotta</b>
Kaelin, M.A., Huebner, W.W., Cordell, R.L. & Szklarczuk, B.	2008	Articolo	STU	6-12	Altre discipline scientifiche				Approfondimento teorico
Karim, M. N.	2011	Articolo	INS		Autoefficacia degli insegnanti	Iran		Studio dimostrativo	Altra evidenza quantitativa
Karri, S. K. R., & Kode, S.	2011	Atti di convegno	INS	Università	Altre discipline scientifiche	India	2008-in corso	Politica a scala	Altra evidenza quantitativa
Kim, S., Chang, M. & Kim, H.	2011	Articolo	STU	Scuola dell'infanzia	Matematica, altre discipline scientifiche	Stati Uniti	2003-2004	Politica a scala	Altra evidenza quantitativa
Kinzie, M. B., Whitaker, S. D., Neesen, K., Kelley, M., Matera, M., & Pianta, R. C.	2006	Articolo	INS, STU	Scuola dell'infanzia	Lingua nazionale	Stati Uniti		Politica a scala	Altra evidenza quantitativa
Klieger, A., Bar-Yossef, N.	2011	Articolo	INS			Israele		Studio dimostrativo	Altra evidenza quantitativa
Kubitskey, B. W., Vath, R.J., Johnson, H. J., Fishman, B. J., Konstantopoulos, S., Park, G. J.	2012	Articolo	INS	9--13	Altre discipline scientifiche	Stati Uniti		Studio dimostrativo	Altra evidenza quantitativa
Kyriakides, L., Creemers, BPM, & Antoniou, P.	2009	Articolo	STU, INS	1-5		Cipro		Studio dimostrativo	Altra evidenza quantitativa
Lee 2008	2008	Articolo		1-5	Scienze e comprensione testo				Disegno sperimentale
LeLoup, J.W. & Schmidt-Rinehart, B.C.	2003	Articolo	INS		Lingua straniera (L2-spagnolo)	Stati Uniti	2003	Studio dimostrativo	Evidenza qualitativa
Love, T.E. & Hildebrand, D,K.	2002	Articolo	STU	Università-Post-Università (Master, dottorato, ecc.)	Altre discipline scientifiche (statistica)	Stati Uniti	1986-2000	Studio dimostrativo	Approfondimento teorico

Tipologia di campione: Studenti (STU), Insegnanti (INS), Scuole (SCU)

**Classificazione documenti**

**APPENDICE A**

<b>Autori</b>	<b>Anno pubblicazione</b>	<b>Tipo di documento</b>	<b>Tipologia di campione</b>	<b>Grado di Scuola</b>	<b>Focus di riferimento</b>	<b>Paese di realizzazione</b>	<b>Anni di realizzazione</b>	<b>Tipo di esperienza</b>	<b>Evidenza prodotta</b>
Lozano, A. S., Padilla, A. M., Sung, H., & Silva, D. M.	2004	Articolo	INS	1-13	Altre discipline scientifiche	Stati Uniti	1997-2000	Politica a scala	Altra evidenza quantitativa
Lumpe, A., Czerniak, C., Haney H. & Belyukova S.	2012	Articolo	INS, STU	1-8	Scienze	Stati Uniti		Studio dimostrativo	Altra evidenza quantitativa
Marek 1991	1991	Articolo	INS, STU	Scuola dell'infanzia; 1-5	Scienze	Stati Uniti		Studio dimostrativo	Evidenza controfattuale non sperimentale
Martin, S. N., & Scantlebury, K.	2009	Articolo		9-13	Altre discipline scientifiche	Stati Uniti		Studio dimostrativo	Evidenza qualitativa
Martin, W., Strother, S., Beglau, M., Bates, L., Reitzes, T., Culp, K.	2010	Articolo	INS, STU	1-5		Stati Uniti	2007-2009		Altra evidenza quantitativa
Masters, J., Magidin De Kramer R., O'Dwyer L., Dash S., Russel M., 2010	2010	Articolo	INS	1-5	Lingua nazionale	Stati Uniti	2007-2009	Studio dimostrativo	Altra evidenza quantitativa
Matsumura, C., Garnier, H.E. & Spybrook, J.	2012	Articolo	INS, STU	4-5	Pratiche didattiche	Stati Uniti	2006-2008	Studio dimostrativo	Disegno sperimentale
Matsumura, LC, Garnier, H.E, Correnti, R., Junker, B. & DiPrima Bickel, D.	2010	Articolo	INS	1-5	Lingua nazionale	Stati Uniti	2005-2008	Studio dimostrativo	Disegno sperimentale
McCutchen 2002	2002	Articolo	INS, STU	Scuola dell'infanzia; 1-5	Lettura, lingua nazionale			Studio dimostrativo	Evidenza controfattuale non sperimentale
McDonough, K.	2006	Articolo	INS		Lingua straniera (L2)	Stati Uniti		Politica a scala	Evidenza qualitativa
McGill-Franzen 1999	1999	Articolo	INS, STU	Scuola dell'infanzia	Lettura, lingua nazionale	Stati Uniti		Studio dimostrativo	Disegno sperimentale
McKenzie, T.L., Stone, E.J., Feldman, H.A., Epping, J.N., Yang, M., Strikmiller, P.K., Lytle, L.A., Parcel, G.S.	2001	Articolo	STU, INS	1-5	Educazione motoria	Stati Uniti	1991-1994	Studio dimostrativo	Disegno sperimentale
Merrill, C., Devine, K. L., Brown, J. W., Brown, R. A.	2010	Articolo	INS, STU	9-13	Matematica	Stati Uniti	2009	Studio dimostrativo	Evidenza qualitativa

Tipologia di campione: Studenti (STU), Insegnanti (INS), Scuole (SCU)

**Classificazione documenti**

**APPENDICE A**

<b>Autori</b>	<b>Anno pubblicazione</b>	<b>Tipo di documento</b>	<b>Tipologia di campione</b>	<b>Grado di Scuola</b>	<b>Focus di riferimento</b>	<b>Paese di realizzazione</b>	<b>Anni di realizzazione</b>	<b>Tipo di esperienza</b>	<b>Evidenza prodotta</b>
Mikami, A. Y., Gregory, A., Allen, J. P., Pianta, R. C., Lun, J.	2011	Articolo	STU, INS	9-13	Relazione studenti-insegnanti	Canada		Studio dimostrativo	Altra evidenza quantitativa
Moyer-Packenham, P. S., Bolyard, J. J., Oh, H., & Cerar, N. I.	2011	Articolo	INS, SCU, STU		Matematica	Stati Uniti			Review/Meta-analisi
Musset, P.	2010	Working paper							Review/Meta-analisi
OECD	2009	Capitolo su volume	INS, SCU, STU						Approfondimento teorico
OECD	2005	Capitolo su volume							Approfondimento teorico
OECD	2011	Capitolo su volume							Approfondimento teorico
Owston, R., Wideman, H., Murphy, J., & Lupshenyuk, D.	2008	Articolo	STU, INS	1-13	Altre discipline scientifiche	Stati Uniti		Studio dimostrativo	Evidenza qualitativa
Pantić, N., Wubbels, T., & Mainhard T.	2011	Articolo	INS			Bosnia e Herzegovina, Croatia, Macedonia, Montenegro, Serbia		Studio dimostrativo	Altra evidenza quantitativa
Peacock, A., & Rawson, B.	2001	Articolo	INS	1-13	Diverse discipline	Sri Lanka e Sud Africa		Studio dimostrativo	Evidenza qualitativa
Pianta, R. C., Mashburn, A. J., Downer, J. T., Hamre, B. K., & Justice, L.	2008	Articolo	INS	Scuola dell'infanzia	Altre discipline scientifiche	Stati Uniti	2004-2005	Politica a scala	Disegno sperimentale
Pinnock, A.	2006	Tesi di dottorato	INS	Università	Altre discipline scientifiche	Giamaica		Studio dimostrativo	Altra evidenza quantitativa
Porter, A.C., Blank, R.K., Smithson, J.L. & Osthoff, E.	2005	Articolo		6-8	Matematica, altre discipline scientifiche	Stati Uniti	2005-	Politica a scala	Disegno sperimentale
Quint. J.	2011	Rassegna di esperienze			Lingua nazionale, Matematica	Stati Uniti			Review/Meta-analisi

Tipologia di campione: Studenti (STU), Insegnanti (INS), Scuole (SCU)

## Classificazione documenti

## APPENDICE A

Autori	Anno pubblicazione	Tipo di documento	Tipologia di campione	Grado di Scuola	Focus di riferimento	Paese di realizzazione	Anni di realizzazione	Tipo di esperienza	Evidenza prodotta
Randel, B., Beesley, A. D., Aphorp, H., Clark, T.F., Wang, X., Cicchinelli, L. F., & Williams, J. M.	2011	rapporto di ricerca	INS, SCU, STU	1-5	Matematica	Stati Uniti	2006-2011	Studio dimostrativo	Disegno sperimentale
Reynolds, A., Ross, S. M., & Rakow, J. H.	2002	Articolo	INS, SCU	6-8		Stati Uniti	non segnato	Politica a scala	Altra evidenza quantitativa
Rivkin, S.G., Hanushek, E.A., & Kain, J.F.	2005	Articolo	STU	1-8	Matematica, Lingua nazionale	Stati Uniti	1993-2000	Studio dimostrativo	Evidenza controfattuale non sperimentale
Roehrig, G. H., Dubosarsky, M., Mason, A., Carlson, S., Murphy, B.	2011	Articolo	INS, STU	1-5	Altre discipline scientifiche	Stati Uniti		Studio dimostrativo	Evidenza qualitativa
Ross, J. A. & Bruce, C. D.	2007	Articolo	INS	6-8	Pratiche didattiche	Canada		Studio dimostrativo	Disegno sperimentale
Ross, J. A., Hogaboam-Gray, A., & Hannay, L.	2001	Articolo	STU, INS	1-5	Altre discipline scientifiche (competenze informatiche)	Canada			Altra evidenza quantitativa
Sailors-Price 2010	2010	Articolo	INS, STU	1-8	Lettura, lingua nazionale				Disegno sperimentale
Santagata, R., Kersting, N., Givvin, K.B. & Stigler, J.W.	2011	Articolo	INS, STU	6-8	Matematica	Stati Uniti	non segnato	Studio dimostrativo	Disegno sperimentale
Santerini, M.	2010	Capitolo su volume	INS		Comportamenti	Italia			Evidenza qualitativa
Schulz, R.A.	2000	Articolo			Lingua straniera (L2)				Approfondimento teorico
Shymansky, J. A., Wang, T., Annetta, L. A., Yore, L.	2012	Articolo	INS, SCU, STU	1-5	Altre discipline scientifiche	Stati Uniti	2000-2005	Studio dimostrativo	Altra evidenza quantitativa
Sinclair, B. B., Naizer, G., Ledbetter, C.	2011	Articolo	INS	1-8	Altre discipline scientifiche	Stati Uniti		Studio dimostrativo	Altra evidenza quantitativa
Slavin, R. & Lake, C.	2007	Rassegna di esperienze	STU	1-5	Matematica	Stati Uniti			Review/Meta-analisi
Slavin, R., Cheung, A., Groff, C., & Lake, C.	2008	Rassegna di esperienze	STU	6-13	Lingua nazionale	Stati Uniti			Review/Meta-analisi

Tipologia di campione: Studenti (STU), Insegnanti (INS), Scuole (SCU)

**Classificazione documenti**

**APPENDICE A**

<b>Autori</b>	<b>Anno pubblicazione</b>	<b>Tipo di documento</b>	<b>Tipologia di campione</b>	<b>Grado di Scuola</b>	<b>Focus di riferimento</b>	<b>Paese di realizzazione</b>	<b>Anni di realizzazione</b>	<b>Tipo di esperienza</b>	<b>Evidenza prodotta</b>
Slavin, R., Lake, C. & Groff, C.	2008	Rassegna di esperienze	STU	6-13	Matematica	Stati Uniti			Review/Meta-analisi
Slavin, R., Lake, C., Chambers, B., Cheung, A. & Davis, S.	2009	Rassegna di esperienze	STU	1-5	Lingua nazionale	Stati Uniti			Review/Meta-analisi
Slavin, R., Lake, C., Cheung, A. & Davis, S.	2009	Rassegna di esperienze	STU	2-5	Lingua nazionale	Stati Uniti			Review/Meta-analisi
Slavin, R., Lake, C., Davis, S., & Madden, N.	2011	Rassegna di esperienze	STU	1-5	Lingua nazionale	Stati Uniti			Review/Meta-analisi
Slavin, R., Lake, C., Hanley, P. & Thurston, A.	2012	Rassegna di esperienze	STU	1-5	Scienze	Stati Uniti			Review/Meta-analisi
Sloan 1993	1993	Articolo	INS, STU	1-5	Matematica, Lingua nazionale, Scienze.	Stati Uniti		Studio dimostrativo	Disegno sperimentale
Taylor, M. Yates, A., Meyer, L. H., Kinsella, P.	2011	Articolo		9-13		Nuova Zelanda		Politica a scala	Altra evidenza quantitativa
Telese, J.	2012	Articolo	INS, STU	6-8	Matematica	Stati Uniti	2005	Studio dimostrativo	Altra evidenza quantitativa
Telese, J.	2008	Articolo							Approfondimento teorico
Tenken 2003	2003	Articolo	INS, STU	1-5	Lettura, lingua nazionale	Stati Uniti		Studio dimostrativo	Disegno sperimentale
Thurston, A., Christie, D., Howe C. J., Tolmie A. & Topping, K.J.	2008	Articolo	STU	1-5	Scienze	Scozia	non segnato	Studio dimostrativo	Evidenza controfattuale non sperimentale
Tschirner, E.	2007	Articolo	INS	9-13	Lingua straniera (L2- tedesco)	Stati Uniti	2005	Studio dimostrativo	Evidenza qualitativa
Vaughn 2011	2011	Articolo	INS	6-8	Lettura, lingua nazionale				Disegno sperimentale
Villegas-Reimers, E.	2003	Rassegna di esperienze	STU	1-13					Review/Meta-analisi - Approfondimento teorico
Weiss, I. R., Pasley, J. D.	2006	Articolo	STU, INS	6-8	Matematica	Stati Uniti			Altra evidenza quantitativa

Tipologia di campione: Studenti (STU), Insegnanti (INS), Scuole (SCU)

## Classificazione documenti

## APPENDICE A

<b>Autori</b>	<b>Anno pubblicazione</b>	<b>Tipo di documento</b>	<b>Tipologia di campione</b>	<b>Grado di Scuola</b>	<b>Focus di riferimento</b>	<b>Paese di realizzazione</b>	<b>Anni di realizzazione</b>	<b>Tipo di esperienza</b>	<b>Evidenza prodotta</b>
Ye He*, K. P. & Teneka, S.	2011	Articolo	INS, STU		Lingua nazionale	Stati Uniti	2007-2008	Studio dimostrativo	Altra evidenza quantitativa
Yoon, K.S., Duncan T., Wen-Yu Lee, S., Scarloss, B. & Shapley, K.L.	2007	Rassegna di esperienze	STU	1-13	Matematica, Lingua nazionale, Scienze.	Stati Uniti	1986-2003		Review/Meta-analisi
Yoon, K.S., Garet. M.S., Birman, B. & Jacobson, R.	2006	rapporto di ricerca	INS	1-13	Matematica, Scienze.	Stati Uniti	2002-2004	Studio dimostrativo	Altra evidenza quantitativa
Zambo, R., Zambo, D.	2008	Articolo	STU, INS	1-13	Matematica	Stati Uniti			Altra evidenza quantitativa

Tipologia di campione: Studenti (STU), Insegnanti (INS), Scuole (SCU)

## APPENDICE B

Griglie di analisi

---

Tabella 1 - Effetti

Study	Study design (RCT or QED)*	Teach/ Stud	Outcomes	Test used	Effect	Effect size	Correction for clustering	Recomputed statistical significance
Arens et al, 2012	RCT	Students	English language acquisition	IDEA Proficiency Test (IPT)		- 0.03	Yes	Not statistically significant
		Teacher	Teacher training and professional experience	Teacher background surveys				
			Classroom practices	Teacher logs and classroom observations				
			Teacher implementation and support received	Teacher interview				
Argentin et al. 2014	RCT	Teacher	Preparation of teaching materials jointly with colleagues. 1-3 times a week ( <i>Ref: less frequently</i> )	Teacher survey	12.3** ITT; 31.5** LATE		Yes (standard errors clustered at the school level)	Statistically significant
			Exchange of point of views on how to teach a particular topic. 1-3 times a week ( <i>Ref: less frequently</i> )	Teacher survey	7.2 ITT; 18.4 LATE		Yes (standard errors clustered at the school level)	Non statistically significant but substantially important
				Teacher survey		-0.31** ITT, -0.78** LATE	Yes (standard errors clustered at the school level)	Statistically significant
				Teacher survey		-0.17* ITT, -0.43* LATE	Yes (standard errors clustered at the school level)	Statistically significant
		Students	Math score	SNV - Italian Mathematics National Assessment test		0.01 ITT 0.03 LATE ATTENTION: RESULTS DIFFER BY AGE OF TEACHER!	Yes (standard errors clustered at the classroom level)	Non significant
			Skipping items (%)	SNV - Italian Mathematics National Assessment test	5.8* ITT 14.6* LATE		Yes (standard errors clustered at the classroom level)	Statistically significant
			Not completing the test (%)	SNV - Italian Mathematics National Assessment test	1.2 ITT 3.1 LATE		Yes (standard errors clustered at the classroom level)	Non significant
			Attitude towards Math, index	SNV questionnaire		0.06* ITT, 0.016* LATE	Yes (standard errors clustered at the classroom level)	Statistically significant
			Perception of curriculum pace	SNV questionnaire	0.08** ITT, 0.018** LATE		Yes (standard errors clustered at the classroom level)	Statistically significant
			Attribution of failures to bad luck	SNV questionnaire	-0.03** ITT, -0.07** LATE		Yes (standard errors clustered at the classroom level)	Statistically significant
			Test Anxiety, index	SNV questionnaire		0.06* ITT, 0.015* LATE	Yes (standard errors clustered at the classroom level)	Statistically significant

\*RCT is a randomized controlled trial; QED is a quasi-experimental design.

Tabella 1 - Effetti

Study	Study design (RCT or QED)*	Teach/ Stud	Outcomes	Test used	Effect	Effect size	Correction for clustering	Recomputed statistical significance
Booth Olson et al, 2012	RCT	Students	Ability to understand, interpret, and write analytical essays	The Assessment of Literary Analysis (on-demand, writing assessment).		0.35 (year 1) 0.67 (year 2)	Yes	Statistically significant
			Mastery of the English language arts content standards: reading	California Standards Test		0.083 (2 year)	Yes	
			Mastery of the English language arts content standards: writing	California Standards Test		0.10 (2 year)	Yes	Statistically significant
		Teacher	Intervention delivery	Quality Checklist			Yes	
		Class	Fidelity of implementation	Classroom observation (only for a subsample of classes)				
Bos et al., 2012	RCT	Students	English language proficiency	California Standards Test for English Language Arts (CST-ELA)		"- 0.01" (7 grade) "0.01" (8 grade) tested in 2010		Not statistically significant
			English language proficiency	California English Language Development Test (CELDT)		0.05 (6 grade) and 0.03 (7 grade) tested in 2009		Not statistically significant
		Teacher	Classroom practice	Sheltered Instruction Observation Protocol (SIOP)		0.0111111111111111		Not statistically significant
Carpenter et al., 1989	RCT	Students	Math Computation	Iowa Test of Basic Skills Level 7 (standardized math test)		0.41	None applied if author did not report significant results	Not significant, but substantively important
			Math Problem solving	Iowa Test of Basic Skills Level 7 (standardized math test)		0.41	None applied if author did not report significant results	Not significant, but substantively important
			Problem solving strategies	Inteview		-	-	-

\*RCT is a randomized controlled trial; QED is a quasi-experimental design.

Tabella 1 - Effetti

Study	Study design (RCT or QED)*	Teach/ Stud	Outcomes	Test used	Effect	Effect size	Correction for clustering	Recomputed statistical significance
Cavalluzzo et al., 2012	RCT	Students	Grade 10 math course enrollment	Math course enrollment records in the post- intervention year		na	Yes	Not statistically significant
			Math achievement	American College Testing (ACT) PLAN: pre-algebra/algebra portion		-9	Yes	Not statistically significant
		Teacher	teacher perceptions of their algebra I approach (hybrid or district curriculum)	Teacher survey				
			Teacher or student centered strategies	Classroom observation (the School Observation Measure and the Algebra I Quality Assessment.)				
Cole, 1992	RCT	Students	Math achievement	Stanford Achievement Test		0.50	Yes	Statistically significant
			Reading achievement	Stanford Achievement Test		0.82	Yes	Statistically significant
			Language achievement	Stanford Achievement Test		0.24	None applied if author did not report significant results	Non statistically significant
Duffy et al., 1986	RCT	Students	Reading comprehension	Gates-MacGinitie Word Reading Subtests		0.00	None applied if author did not report significant results	Non statistically significant
Garet et al., 2008	RCT	Teachers (ITT)	Teacher Knowledge	Reading content and Practiced Survey	-	a. 0.37* (0.18 f.u.y) b. 0.38* (0.07 f.u.y)	yes (using multilevel model)	Statistically significant
			Word level subscale		-	a. 0.35* (0.17 f.u.y) b. 0.39* (0.18 f.u.y)		Statistically significant
			Meaning level subscale		-	a. 0.21 (0.01 f.u.y) b. 0.26 (-0.11 f.u.y)		Not significant, but substantively important
			Use of explicit instructions		Trained Observer	-		a. 0.33* (0.09 f.u.y) b. 0.53*(-0.03 f.u.y.)
		Students (ITT)	Student reading test scores	District assessment	-	a. 0.08 (0.10 f.u.y) b. 0.03 (0.01 f.u.y.)		Non statistically significant

\*RCT is a randomized controlled trial; QED is a quasi-experimental design.

Tabella 1 - Effetti

Study	Study design (RCT or QED)*	Teach/ Stud	Outcomes	Test used	Effect	Effect size	Correction for clustering	Recomputed statistical significance
Garet et al, 2011	RCT	Teachers	Teacher Knowledge	Constructed specifically for the study	0.05	0.05	yes (using multilevel model)	Non statistically significant
			Common Knowledge subsc.		-0.29	-0.21		Non statistically significant
			Specialized Knowledge subsc.		0.41	0.36		Not significant, but substantively important
		Students	Rational Number Achievement	NWEA rational number test (constructed for the study)	-0.07	-0.01		Non statistically significant
			Fraction and decimal subscale		-0.21	-0.01		Non statistically significant
			Ratio and proportion subscale		0.14	0.01		Non statistically significant
Gersten et al., 2010	RCT	Teachers	Practice (external observer)	Reading comprehension		0.86**	yes (using multilevel model)	Statistically significant
				Vocabulary		0.58**		Statistically significant
			Knowledge	Reading Vocabulary		0.73*		Statistically significant
				Content Knowledge for Teaching Reading		0.32		Not significant, but substantively important
		Students	Reading vocabulary	Diebels		0.21		Non statistically significant
			Oral vocabulary	Diebels		0.44		Not significant, but substantively important
			Passage comprehension	WDRB		0.13		Non statistically significant
			Letter-word identification	WDRB		0.21		Non statistically significant
			Word attack	WDRB		0.13		Non statistically significant
			Other reading fluency	WDRB		0.23		Non statistically significant

\*RCT is a randomized controlled trial; QED is a quasi-experimental design.

Tabella 1 - Effetti

Study	Study design (RCT or QED)*	Teach/ Stud	Outcomes	Test used	Effect	Effect size	Correction for clustering	Recomputed statistical significance
Glazerman et al., 2010	RCT	Student	Reading achievement	Standardized reading test		0.11 in the third year (2-year intervention)	Yes	Statistically significant
		Student	Math achievement	Standardized math test		0.20. in the third year (2-year intervention)	Yes	Statistically significant
		Teacher	background of the study teachers, receipt of induction services and alternative support services, teacher attitudes, and mobility patterns was administered six times over four years.	Teacher survey			Yes	
			Teaching practice	Classroom observation			Yes	
Greenleaf et al., 2011	RCT	Teachers	Change in classroom practices	Survey (teacher self reported)		0.61**		average effect size on the 14 subconstruct with statistically significant changes
		Students (subgroup of English Language Learner: ELL)	Biology and reading comprehension	Opportunity To Learn (OTL)		0.12 (0.3 ELL*)	yes (using multilevel model)	Statistically significant only for ELL
			English Language Arts	California standardized test scores (CST)		0.23** (ELL 0,17*)		Statistically significant
			Reading comprehension			0.24** (ELL 0,21**)		Statistically significant
			Biology			0.28** (ELL 0,23**)		Statistically significant
Grigg 2012	QED	Teacher	changes in teaching practice	classroom observations	The lessons from the Immersion schools were nearly twice as likely to show evidence of any kind of scientific inquiry across all 3 years of the study (odds ratio = 1.84). This contrast was strongest in the 1st year (odds ratio = 2.51).		cluster-adjusted	statistically significance

\*RCT is a randomized controlled trial; QED is a quasi-experimental design.

Tabella 1 - Effetti

Study	Study design (RCT or QED)*	Teach/ Stud	Outcomes	Test used	Effect	Effect size	Correction for clustering	Recomputed statistical significance
HellerJ_2012	RCT	Teacher/student	Student content knowledge of force and motion; Student content knowledge of physical science more generally; Teacher content knowledge of force and motion; Confidence in ability to teach	Assessing Teacher Learning About Science Teaching (ATLAST) Test of Force and Motion for Students (Smith and Banilower 2006a, 2006b); California Standards Test reporting clusters on motion (8 items) and forces, density, and buoyancy (13 items) (California Department of Education, 2011b); ATLAST Test of Force and Motion for Teachers (Smith and Banilower 2006a, 2006b); Teacher survey about confidence		ATLAS for student 0.11 and California Standard test 0.03(students' content knowledge ); ATLAS for ELL 0.31 and California standard test 0.09; ATLAS for teachers 0.38** (teacher science content knowledge); confidence in ability to teach force and motion 0.49**	adjusted for clustering	not statistically significant for student achievement; statistically significant for teacher outcomes
Kim et al., 2011	RCT	Students	Reading comprehension	Assessment of Literacy Analysis		0.35***	yes (using multilevel model). Student nested in 101 class for ALA and 95 for CST	Statistically significant
				CST total		0.068*		Statistically significant
				CST reading subset		0.046		Non statistically significant
			Writing skills	CST writing subset		0.087		Non statistically significant
Lee et al., 2008	RCT	Students	Science (Comparison PRE-POST)		8.65***	-		
			Math (Treated-Control)	National Stand Test		0.47**	yes (using multilevel model)	Statistically significant
Marek & Methven, 1991	QE	Students	Conservation reasoning	Piagetian cognitive tasks		0.39	Yes	Statistically significant
Masters et al., 2010	RCT	Teachers	Vocabulary knowledge	Developed specifically for the study	ANCOVA (partial eta-squared)	0.038	No clustering problem, because the results are on teachers and randomization was at the teacher level	Non statistically significant
			Reading comprehension knowl			0.172***		Statistically significant
			Writing knowledge			0.248***		Statistically significant
			ELA knowledge			0.170***		Statistically significant
			Vocabulary practices	Self-reported		0.150***		Statistically significant
			Reading comprehension practices			0.068*		Statistically significant
			Writing practices			0.142***		Statistically significant

\*RCT is a randomized controlled trial; QED is a quasi-experimental design.

Tabella 1 - Effetti

Study	Study design (RCT or QED)*	Teach/ Stud	Outcomes	Test used	Effect	Effect size	Correction for clustering	Recomputed statistical significance
Matsamura et al., 2010	RCT	Teachers	Classroom practices	Self-reported				Categorical variables: results not comparable to the others here presented
		Students	Reading comprehension skills (Effect in TASK point)	TASK	21.02 (59 ELL)	0.17 (0.51**ELL)	yes (using multilevel model)	Statistically significant for ELL
Matsamura et al., 2012	RCT	Teacher/student	Classroom practices: quality of text discussion	Quality of classroom text discussion, teacher survey reports, literacy-coach interviews		0.89		statistically significant
McCutchen et al., 2002	QED	Students	Word reading	Gates-MacGinitie Word Reading Subtests		0.39	No	Statistically significant
McGill-Franzen et al., 1999	RCT	Students	Early literacy skills	Concepts about print		1.11	Yes	Statistically significant
		Students	Early literacy and writing skills	Letter identification		0.69	Yes	Statistically significant
		Students	Early literacy and writing skills	Writing vocabulary		0.32	Yes	Statistically significant
		Students	Early literacy and writing skills	Ohio word test		0.66	Yes	Not significant, but substantively important
		Students	Early literacy and writing skills	Hearing sounds in words		0.97	Yes	Not significant, but substantively important
		Students	Receptive language skills	Peabody Picture Vocabulary Test		0.12	None applied if author did not report significant results	Statistically significant
Pianta 2008	RCT	Teacher	teacher-child interactions	Teachers survey, PALS, CLASS	teacher sensitivity 0.07*; Instruction learning formats 0.07*; Language modeling 0.09**			For three dimensions of interaction quality – Teacher Sensitivity, Instructional Learning Formats, and Language Modeling – the rates of change were significantly different between teachers who participated in the Consultation condition and teachers who participated in the Web Only condition
Randel 2011	RCT	student	math achievement	mathematics subtest of the Colorado Student Assessment Program (CSAP)		0.01	not adjusted for clustering	Non statistically significant

\*RCT is a randomized controlled trial; QED is a quasi-experimental design.

Tabella 1 - Effetti

Study	Study design (RCT or QED)*	Teach/ Stud	Outcomes	Test used	Effect	Effect size	Correction for clustering	Recomputed statistical significance
Sailors & Price, 2010	QED	Teachers	Opportunity to engage	Comprehension Instruction Observation Protocol System (CIOPS)		0.78*		Statistically significant
			Constructed explanations			0.64		Not significant, but substantively important
		Students	Reading comprehension	Group Reading Assessment and Diagnostic Evaluation (GRADE)		0.33**	yes (using multilevel model)	Statistically significant
Santagata2011	RCT	Teacher	teacher knowledge; classroom practice; students' learning;	Developed specifically for the study for the teachers and District-wide Quarterly Assessment and California Standard test for the students	It not observe a treatment effect of the PD program on student performance on the Quarterly Assessment ( $\beta = -.16$ ; Model 1).It did observe a statistically significant treatment-teacher content knowledge interaction for Model 2 ( $\beta = 11.37$ ). PD effects on student learning differed based on the level of teacher content knowledge.	teacher knowledge: full scale 0.095; 0.057 for the pedagogical content knowledge subscale, 0.076 for the content knowledge subscale		
Saxe et al., 2001	QED	Students	Fraction concept	29-item, 40-minute times measure developed by the authors		2.39	No	Statistically significant
		Students	Fraction computation	29-item, 40-minute times measure developed by the authors		-0.53	No	Not significant, but substantively important

\*RCT is a randomized controlled trial; QED is a quasi-experimental design.

Tabella 1 - Effetti

Study	Study design (RCT or QED)*	Teach/ Stud	Outcomes	Test used	Effect	Effect size	Correction for clustering	Recomputed statistical significance
Sloan, 1993	RCT	Students	Reading	Comprehensive Test of Basic Skills		0.68	Yes	Not significant, but substantively important
		Students	Math	Comprehensive Test of Basic Skills		0.26	None applied if author did not report significant results	Not significant, but substantively important
			Science	Comprehensive Test of Basic Skills		0.63	Yes	Not significant, but substantively important
			Self-esteem					
		Teachers		Classroom observation				
		Class	classroom environment					
Tienken, 2003	RCT(with group equivalence problems)	Students	Narrative writing	Content/Organization score on a standardized writing test administered as part of New Jersey's Elementary School Proficiency Assessment		0.41	Yes	Not significant, but substantively important
		Teacher		Reflective logs				
		Class		Classroom observation				
Vaughn et al., 2011	RCT	Students	Reading comprehension	Gates-MacGinitie	1.21	0.12* (0.36)	yes (using multilevel model)	Statistically significant for the full sample Not significant, but substantively important for the struggling students

\*RCT is a randomized controlled trial; QED is a quasi-experimental design.

Tabella 2 - Caratteristiche

Study	Name or Type of PD	Subject	Target	Level of intervention	Grade	Form of PD	Content	*Kind of PD	Intensity	Participation
Arens et al, 2012	PD provides K–5 teachers with learning opportunities to acquire the knowledge and skills to support ELL students in language and literacy learning	English language	English language learner (ELL); students who speak Spanish as their primary language	Teacher/classroom	Primary school	During Year 1, the training year, the publisher provided PD for teachers in the intervention group on the use of OWE in their classrooms and intervention teachers were trained in the use of RISE strategies. In year 2, teachers in the intervention group were asked to use OWE each day for at least 30 minutes per class and implement the RISE techniques fully in their classrooms. During their training, teachers discuss teaching practices with their peers, view videotaped lessons, and read articles about curriculum. Additional classroom support is provided by a facilitator who performs instructional demonstrations and coaching	The PD program, entitled Responsive Instruction for Success in English (RISE), is a PD program designed to meet the needs of K–5 teachers by providing them with sustained adult learning opportunities to acquire the knowledge and skills to support ELL students in language and literacy learning. The classroom materials are entitled On Our Way to English (OWE). OWE is a comprehensive English curriculum for elementary classrooms (grades K–5) developed to provide ELL students with simultaneous access to English oral language development, comprehensive literacy instruction, and standards-based content area information in science and social studies	Kennedy #2	Coordinators (who received a 3-day training) deliver 8 RISE instructional modules to teachers in the form of interactive workshops. Teachers also received a half-day OWE training session.	
Argentin et al. 2014	M@t.abel+	Math	Southern Italian Regions	Teachers	6th to 8th	Blended training course on teaching skills	It aims at increasing lower secondary school math achievement, providing teachers with alternative solutions and methods for presenting traditional contents. The main idea is that students should be engaged in solving real life problems through mathematical tools and concepts. There is a repository of teaching materials facing different curriculum math concepts by adopting a problem-solving perspective. Teachers are required to use at least four of these teaching materials (precisely one per major math content area) in their classrooms and to report on the experience to their tutor and peers through a structured diary. Moreover, the program encourages a virtual community of teachers to exchange views through on-line forums and discussion groups, also from home	Kennedy #2	26 hours of formal training and 30 hours of on-line training along 1 school year	49% completed the training but only 39% were compliant to the treatment

\*Kennedy content area

Tabella 2 - Caratteristiche

Study	Name or Type of PD	Subject	Target	Level of intervention	Grade	Form of PD	Content	*Kind of PD	Intensity	Participation
Booth Olson et al, 2012	PD supports a cognitive strategies reading/writing intervention for mainstreamed English learners (ELs) in the secondary grades.	English language arts (Reading and writing)	Students are classified as being at the poverty level, and English learners		Secondary (6th to 12th)	There are three core components of the Pathway Project: (1) training in the use of the cognitive strategies tool kit and curriculum materials, (2) intervention activities focused on the revision of the pretest on-demand writing assessment into a multiple draft essay, and (3) coaching from a more experienced, veteran teacher previously trained in the Pathway Project on how to integrate a cognitive strategies approach into the existing English language arts curriculum	Pathway Project, a cognitive strategies approach to teaching interpretive reading and analytical writing, or to a control condition involving typical district training focusing on teaching content from the textbook. Pathway teachers learned how to use an on-demand writing assessment to help mainstreamed English learners understand, interpret, and write analytical essays. Secondary English teachers are trained to improve mainstreamed ELs' interpretive reading and text-based analytical writing abilities by: (1) using a cognitive strategies approach to reading and writing instruction, (2) instructing students to revise a pretest on-demand writing assessment into multiple draft essays, and (3) receiving ongoing support from experienced Pathway teachers who serve as coaches to the teachers in the experimental condition		46 hours of training (via six 6-hour released days during the school year and five 2-hour afterschool sessions). This was in addition to 26 hours of typical professional development that all district secondary English teachers received	

\*Kennedy content area

Tabella 2 - Caratteristiche

Study	Name or Type of PD	Subject	Target	Level of intervention	Grade	Form of PD	Content	*Kind of PD	Intensity	Participation
Bos et al., 2012	PD supports a cognitive strategies reading/writing intervention for English learners (ELs) a7 6, 7, 8 grades.	English language	Middle school English language learners	Classroom	Middle school	The summer institutes consist of 7 days of PD group sessions to provide a foundation for using new tools and processes for the academic and linguistic development of adolescent English language learner students. To promote continuity across school years, three days are offered at the end of a school year (June/July) and four days are offered before the start of the next (August/September). Four to six cycles of individualized coaching are offered to teachers participating in QTEL each year. Coaches help teachers develop academically and linguistically rigorous lessons that implement QTEL principles, tools, and processes. These coaching cycles consist of a one-on-one lesson design meeting, an observation of the lesson's implementation, and a debriefing. The collaborative lesson design meetings, a series of monthly planning sessions, are held at the school sites and facilitated by QTEL staff to provide support for QTEL implementation	Quality Teaching for English Learners (QTEL), developed by WestEd, is an approach to improving the teaching of English language learner students at the secondary level. These principles permeate the three core components of QTEL: summer institutes, individualized teacher coaching, and collaborative lesson design meetings	Kennedy #2	7 days of group sessions + 4/6 cycles of individualized coaching + monthly collaborative lesson meetings	
Carpenter et al., 1989	Cognitively guided instruction	Math		Teacher	Elementary (1th)	Four months (workshop and one follow-up meeting)	How students learn math, relationships between math problems and how students process to solve them, research on math acquisition, examination of curricula, how materials affect teaching, planning instruction	Kennedy #4	83 hours over four months	

\*Kennedy content area

Tabella 2 - Caratteristiche

Study	Name or Type of PD	Subject	Target	Level of intervention	Grade	Form of PD	Content	*Kind of PD	Intensity	Participation
Cavalluzzo et al., 2012	Kentucky Virtual Schools' hybrid program for algebra I, an intervention combining online instruction with face-to-face classroom teaching to increase student learning and achievement in grade 9 algebra I, a course required by Kentucky for high school graduation	Math (algebra I)	High school students whose classroom teachers were inexperienced or less successful than desired in specific subject areas	Classroom	High school (9th)	Teachers engage in ongoing PD—through online content used in face-to-face and distance settings—to learn how to teach an algebra course structured on research-based strategies in a hybrid classroom. The PD experience for hybrid algebra I, which begins the summer before the school intervention year, is guided by instruction specialists and uses an online PD program for algebra I teachers, as well as other resources. It then moves to a distance format, with teachers meeting online in guided weekly sessions designed to increase their conceptual understanding of math content and pedagogy. During the school year, monthly online sessions allow teachers to share their successes and challenges to improve their understanding of, and ability to implement, recommended instruction practices. Topics for monthly discussions are selected by the instructional specialists and informed by site visits to schools, where the instruction specialists observe implementation and help guide improvement	Ongoing professional development provides training on hybrid algebra I instructional strategies, such as teacher modeling to reinforce student problem solving, reasoning, and communication through reading, writing, talking, and technology and manipulatives to acquire knowledge and skills	Kennedy #2	One-year ongoing PD	Only about 23% fully completed all assignments
Cole, 1992	Mississippi Teacher Assessment Instruments staff development	Math, Reading, English/language arts		Teacher/classroom (implementation of training materials in class)	Elementary (4th)	Eight three-hour sessions over a two-month period with follow-up observational visits throughout the year, plus two half day follow-up conferences	Modeling of the 14 Mississippi Teacher Assessment Instrument teacher (pedagogical) behavior competencies (for example, planning instruction to achieve selected objectives, organizing instruction to take into account individual differences among learners, and obtaining and using information about the needs and progress of individual learners)	Kennedy #1	40+ hours over a year	

\*Kennedy content area

Tabella 2 - Caratteristiche

Study	Name or Type of PD	Subject	Target	Level of intervention	Grade	Form of PD	Content	*Kind of PD	Intensity	Participation
Duffy et al., 1986	Incorporating explicit verbal explanations during reading instruction	Reading, English/lang uage arts	Low-achieving readers	Teacher	Elementary (5th )	Five two-hour sessions	How to recast teacher skill at prescriptive basal text techniques into strategies for helping students be better readers when removing blockages to meanings; how to make explicit statements about the reading ills being taught; how to organize these statements for presentation to students	Kennedy #2	10 hours over four months	
Garet et al, 2011	Middle school Mathematic PD	Math	Middle school math teachers (eligible schools should have at least 1/3 of students eligible for free or reduced-price lunch)	Classroom	Middle 7th	Instructional: summer seminar + a series of one-day follow-up during the year + coaching visits	PD promoted the teachers' knowledge of rational number topics, including specialized mathematics knowledge that may be useful for teaching these topic. Rational numbers are considered an essential foundation for algebra	Kennedy #3	114 hours: 68 in the 1st y.; 46 in the 2nd y	Average 68% of the intended dosage
Garet et al., 2008	Early Reading PD Interventions	Reading	Students in high poverty schools	Classroom	Elementary 2th	Two treatment groups: A. Instructional B. Instructional + coaching	PD promoted the use of three specific classroom practice: explicit instruction, guiding students in independent practice of reading activities and differentiating instruction to meet individual students' needs. The practice could be connectcd to the core reading program used in the district	Kennedy #2	A. 48 hours B. 48 hours + 60 hours coaching	Average 78% of hours provided
Gersten et al., 2010	Teacher Study Group	Reading	Students in high poverty schools	Classroom	Elementary 1th	Instructional sessions + discussion problem solving groups + applied learning activities (research lessons)	PD for teachers in Reading First Schools. TSG goal was to help teachers begin to think about and ultimately to use research-based instructional concepts in their classrooms by integrating the TSG content into their existing curriculum. TSG promoted also collaborative interaction and collective participation among teachers	Kennedy #2: Focused on teaching behaviour applying to a particular subject	20 hours: 16 session of 75 minutes each	Not specified

\*Kennedy content area

Tabella 2 - Caratteristiche

Study	Name or Type of PD	Subject	Target	Level of intervention	Grade	Form of PD	Content	*Kind of PD	Intensity	Participation
Glazerman et al., 2010	Comprehensive teacher induction to support beginning teachers with a more comprehensive program improves teacher and student outcomes	None	Beginning teacher	Teacher	Elementary school - grades K-6	Beginning teachers were offered monthly professional development sessions, opportunities to observe veteran teachers, and an end-of-year colloquium	Beginning teachers in treatment schools were then assigned to a full-time mentor with a 12 to 1 ratio. Mentors received ongoing training and a curriculum of materials to support the teachers' development. Beginning teachers were offered opportunities to observe veteran teachers, and an end-of-year colloquium	Kennedy #3	In 10 of the 17 districts, the services were offered to treatment schools for one year only. In the remaining 7 districts, services were offered to treatment schools for two years	
Greenleaf et al., 2011	Reading Apprenticeship	Science and Reading	Low-achieving high school students	Classroom	High school	Instructional sessions + follow-up session during school year + actual use of pedagogical tools + participation in school and cross-site learning networks	Two years PD aimed at integrating Literacy apprenticeships into science teaching. It incorporates research-based instructional approach that have been shown to improve adolescent literacy levels in science curriculum and text. PD emphasizes the development of pedagogical content knowledge	Kennedy #3	108 hours: 18 days of PD (10 spread during year + 5 days summer institute + 3 days follow-up)	Not specified
Grigg 2012	1) Science Immersion , 2° FOSS (Full Option Sciences System	Science	LAUSD, one of the country's poorest-performing school districts in elementary science	Teachers	4th and 5 th	Intensive summer training workshop	The Immersion curriculum for fourth grade is a unit called "Rot it Right" that covers the life science standards. The fifth-grade Immersion curriculum is a unit called "Weather Forces and Prediction" that covers the earth science standards. FOSS covers earth, life, and physical science standars. Immersion required teachers to facilitate students' open questioning (open inquiry), whereas the FOSS curriculum was more structured (guided inquiry). Immersion required the classroom teachers to collect and prepare instructional materials on their own, whereas FOSS consisted of prefabricated "FOSS kits	Kennedy #3	Science immersion: 5-day professional development workshop; FOSS took three 1-day professional development workshops, 1 day per science content area	

\*Kennedy content area

Tabella 2 - Caratteristiche

Study	Name or Type of PD	Subject	Target	Level of intervention	Grade	Form of PD	Content	*Kind of PD	Intensity	Participation
HellerJ_2012_L12	Making Sense of Science	Science	urban districts, ELL	Teacher	8th	Workshops were implemented regionally, with local facilitators leading the course for local teachers at each of the six research sites	Making Sense of SCIENCETM is designed to move teachers through learning about key science concepts, literacy supports, classroom practices, and students' science ideas. Courses have four main components: 1) Hands-on science investigations engage teachers in core content dilemmas described in accompanying written teaching cases. 2) Language and literacy activities are intended to teach teachers how to more effectively support students' science reading and discussion skills; help students make sense of the science; and help students, particularly English language learners, develop their academic language proficiency. 3) Case discussions engage teachers in examining detailed instructional scenarios. 4) Classroom connections provide opportunities for teachers to read about, reflect on, and discuss key science and literacy concepts and consider how these concepts pertain to their own work with students	Kennedy #3	Teachers in the intervention group received a 24-hour Making Sense of SCIENCETM professional development course on force and motion in summer 2009, over a period of five days	
Kim et al., 2011	Pathway Project	Writing (Literacy)	Latino English Learner	Classroom	Middle and High 6th to 12th	Instructional + coaching to integrate analytical writing strategies into English Learner Arts (ELA) curriculum.	PD aimed at developing cognitive strategies approach to analytical writing. The cognitive strategy is grounded in a wide body of research on what experienced readers and writers do when they construct meaning from and with text	Kennedy #4	46 hours: 6 full day sessions (6 hours each) and 5 after school sessions (2 hours)	Not specified
Lee et al., 2008	Science and Literacy for ELL	Literacy and Science	English Language Learners (ELL)	Classroom	Elementary 3th	Workshops (before and during school year) + teaching materials	PD aimed at promoting science achievement for culturally and linguistically diverse elementary students. The intervention is focused on integrating English Language and literacy development as a part of science instruction with ELL students	Kennedy #3	30 hours: 5 full day sessions (6 hours each)	Not specified

\*Kennedy content area

Tabella 2 - Caratteristiche

Study	Name or Type of PD	Subject	Target	Level of intervention	Grade	Form of PD	Content	*Kind of PD	Intensity	Participation
Marek & Methven, 1991	Utilizing the Learning Cycle in Elementary School Science	Science	Science teachers		Elementary (K-3rd, 5th )	Four weeks summer workshop	How to develop a curriculum (learning cycles) that represents science, allows students to experience science as a search for knowledge, and is compatible with their students' learning abilities.	Kennedy #2	100 hours over four weeks	
Masters et al., 2010	Online professional development	Language art (Reading and Writing)	No particular target	Classroom	Elementary 4th	Online workshops (each containing: orientation session, readings, activities and discussions).	Online PD aimed at improving teacher's knowledge and instructional practice for 4th grade English language art	Kennedy #2: Focused on teaching behaviour applying to a particular subject	105 hours: 3 workshops lasting 7 weeks each. Requiring 4 to 6 hours participation per week	46% of the recruited completed all the requirements.
Matsamura et al., 2010	Coaching in school with high teacher mobility	Literacy	ELL in low-income, urban schools with high teacher mobility	Teacher	Elementary 4th and 5th	Workshops and coaching	PD for coaches (to work effectively with teachers) and for teachers who were hired to replace teachers that left the school or grade during the second year of a 2 year PD program. PD aimed at building knowledge about effective reading, planning and reflecting on instruction (interaction between coaches and teachers), differentiating instruction for students	Kennedy #2	40 hours: 3 days (18 hours) for coaches and teachers + coaching hours for teachers (at least a weekly individual or small group meeting)	Not specified
Matsamura et al., 2012	Comprehensive literacy - coaching program (CFC content-focused coaching)	literacy	ELL in low-income, urban schools	Teacher	Elementary 4th and 5th	Workshops and coaching	PD provides intensive professional development to literacy coaches aimed at increasing their pedagogical knowledge and skills at teaching a curricula domain, in addition to ensuring their ability to work effectively with teachers. CFC-trained coaches, in turn, return to their school and work with teachers in grade-level teams (i.e., in professional learning groups) and individually in their classroom to improve the quality of instruction	Kennedy #2	Not specified	Not specified

\*Kennedy content area

Tabella 2 - Caratteristiche

Study	Name or Type of PD	Subject	Target	Level of intervention	Grade	Form of PD	Content	*Kind of PD	Intensity	Participation
McCutchen et al., 2002	PD that supports phonological skills instruction	Reading, English/lang uage arts		Teacher	Elementary K-1th	Two-week summer institute plus three follow-up meetings; informal interactions and classroom visits with support	Deepening teachers' understanding of phonology, phonological awareness, analysis of sounds, development of phonological awareness in children, children's mistakes revealing underlying conception of phonemics	Kennedy #4	About 100 hours over 10 months	
McGill-Franzen et al., 1999	PD that supports teachers's instructional routines involving the book collection	Reading, English/lang uage arts			Elementary (kindergarten)	Three whole-day sessions and seven two-hour follow-up sessions	Physical design of the classroom; effective book displays; importance of reading aloud to children; environmental print; author, genre, and content themes created with the book collection; small-group lessons using teacher-made materials based on books read	Kennedy #3	About 30 hours over six months	
Pianta2008	MyTeachingPartner	Teacher-child interaction	School with children income below Federal guidelines for poverty	Teacher	Pre-kindergarten	Web-based system of professional development	The MyTeachingPartner approach to PD includes two components: (a) access to video exemplars of high-quality teacher-child interactions tied to specific dimensions of the CLASS and (b) a consultation process that provides regular, multi-modal, ongoing, targeted feedback to pre k teachers through a standardized protocol that focuses on specific dimensions of teachers' emotional, organizational, and instructional interactions with students, again using theCLASSas the basis for a common, validated understanding of teacher behavior	Kennedy 1	Not specified	Not specified

\*Kennedy content area

Tabella 2 - Caratteristiche

Study	Name or Type of PD	Subject	Target	Level of intervention	Grade	Form of PD	Content	*Kind of PD	Intensity	Participation
Randel 2011	CASL (classroom assessment for student learning)	Math	Public schools in Colorado	teacher	4yh and 5th	CASL is typically implemented via teacher learning teams, in which teachers meet regularly to discuss and reflect on the content of the textbooks and DVDs and to share their experiences applying the program in their classrooms	CASL is a professional development program on classroom and formative assessment published by the Assessment Training Institute of Pearson Education. The CASL program includes a textbook, DVDs, ancillary books, and an implementation handbook, all of which are used to train teachers to conduct classroom assessments that are appropriate for, and aligned with, their learning targets	Kennedy #2	Formal training involved an introductory videoconference and access to district staff who were trained as facilitators. Teachers were asked to form learning teams to study the CASL materials during the 2007–08 school year and to fully implement the CASL practices in 2008–09	The average amount of the time teachers reported spending on CASL training was 31 hours, compared with the 60 hours recommended by the program's developer. approximately 63% of teachers attended the recommended nine learning team meetings called for by CASL
Sailors & Price, 2010	PD that supports the teaching of cognitive reading strategy instruction	Reading	Elementary and middle schools with low income students in Texas	Teacher	1st through 8th	2 models of professional development. One group attended a traditional 2-day summer in-service; the second attended the workshop and received classroom-based support from a reading coach	PD aimed at the improvement of instructional comprehension practices of teachers in regular education classrooms. Two models of PD on the instructional comprehension practices of elementary and middle school teachers and the reading achievement of their students to determine if classroom-based professional development (coaching) proved to be a more effective model PD than a traditional model of professional development (workshop)	Kennedy #2: Focused on teaching behaviour applying to a particular subject	18 hours: 2 days workshop (controls) versus 2 days workshop and 2 years of coaching (treated): 329 minutes per teacher on average in one year	90% completed the fist year

\*Kennedy content area

Tabella 2 - Caratteristiche

Study	Name or Type of PD	Subject	Target	Level of intervention	Grade	Form of PD	Content	*Kind of PD	Intensity	Participation
Santagata2011	PD programm	Math	The district included five low-performing (i.e., having a poverty rate of 50% or higher) inner-city middle schools	Teachers	6th to 8th	Workshop, teachers watched selected lessons, spent a full day on mathematical content area alternating between online tasks and face to-face discussions	Team met monthly with the researchers to plan and videotape the mathematics lessons to be incorporated in the online video-based modules, discuss PD program implementation setting and schedule, and maximize the responsiveness of the program to the specific needs of the students attending the participating schools. The program consisted of three modules, each targeting a key content area (i.e., fractions, ratio and proportion, and expressions and equations), and the respective core concepts. Teachers met in groups of 8 to 10 led by one or both of two facilitators, each having a strong background in mathematics and several years of teaching experience. Within each key content area, teacher learning was structured into three phases: Content Exploration, Lesson Analysis, Link to Practice	Kennedy #3		
Saxe et al., 2001	PD that supports teachers's innovative use of good curriculum materials	Math			Elementary 4th-5th	A weeklong summer workshop with 13 follow-up meetings	Teacher knowledge of mathematics (particularly fractions), teacher knowledge of how students learn mathematics and fractions, and teacher understanding of student motivation in math	Kennedy #4	About 60 hours over six and a half months	
Sloan, 1993	PD that supports teachers's instructional and questioning behaviors	Math, Science, Reading, English/lang uage arts			Elementary 4th-5th	Summer sessions and seven follow-up meetings	Use of instructional and questioning strategies associated with Direct Instruction and Hunter's (1984) Seven Steps of the Teaching Act (for example, anticipatory set, objective and purpose, instructional input, modeling, checking for guidance)	Kennedy #1	About five hours over two months	

\*Kennedy content area

Tabella 2 - Caratteristiche

Study	Name or Type of PD	Subject	Target	Level of intervention	Grade	Form of PD	Content	*Kind of PD	Intensity	Participation
Tienken, 2003	PD that supports environmental model in writing instruction to improve student	Reading, English/lang uage arts			Elementary 4th	Eight one-hour sessions with six follow-up conferences	How to provide instruction to students in the use of the criteria in the New Jersey Registered Holistic Scoring Rubric and a set of high-order reflective questions as self-assessment and reflection devices when composing, revising, and editing narrative essays	Kennedy #3	14 hours over three and a half months	
Vaughn et al., 2011	Collaborative Strategic Reading and metacognitive strategic learning	Reading	Middle schools	Teacher/classroom	Middle 7th and 8th grade	Teachers were provided an initial 3 days training + during year meeting + coaching during CSR session.	The PD focused on: (a) an overview of the study; (b) a carefull description of an experimental study, including the importance of adhering to "business as usual" in control classes and implementing instruc- tional practices in treatment classes; (c) critical features of the intervention practices and how to teach them to students; and (d) how to use collaborative groups within the CSR model. Cooperative learning practices while implementing comprehension strategies in the context of reading are also a critical component of CSR	Kennedy #2	70 hours: 3 days (18 hours) + 3 session during the school year (1.5 each) = 22.5 total + in-class coaching during the 36 (avg.) CSR teaching sessions (50 minutes each)	

\*Kennedy content area

Tabella 3 - Validità

Study	N. of district	N. of schools	N. of teachers	N. of students	Attrition	Randomiz Level	Internal validity	External validity
Arens et al, 2012	13 districts in three Central Region states (Colorado, Kansas, and Nebraska—the states in the REL Central region with the highest enrollments of ELL students)	Schools in the Central Region states with the largest percentages of Spanish-speaking ELL students in the elementary grades were recruited and randomly assigned using a 2:1 ratio to the intervention group (34 schools) or the control group (18 schools).	203 teachers (136 treated and 67 control). In Year 2 19 treated and controls were added.	2612: 1761 treated and 851 controls	54.4% treated, 40.3% controls	School/teacher (site coordinators were asked to identify all teachers at their schools who taught English language development to native Spanish-speaking ELL students in grades 2– 5. At schools in which ELL students did not receive English language development instruction, all literacy teachers who taught native Spanish-speaking ELL students in grades 2–5 were identified as study teachers).	53 schools were randomized, but one intervention school dropped out of the study immediately after randomization. Most teachers in the intervention group received their RISE training from a publisher-trained site coordinator rather than directly from the publisher. Results might have differed had training been provided by the publisher.	This study relied on a volunteer sample of schools that may not be representative of schools across the United States.
Argentin et al. 2014	15 (italian provinces), in 4 regions	174 schools (54 control; 120 treatment, final sample)	581 (172 controls; 409 treatment) (final sample)	about 11000	minimal (8 classes over 581 did not gave access to field)	The schools in each stratum were randomly assigned to the experimental conditions. Then, random assignment of the teacher's class in which to experiment Matabel.	RCT with low level of attrition but high level of non-compliance in the treatment group. Treated and control equivalent on a large set of characteristics at student-teacher-school level	Generalizability limited to the 15 provinces (observables); self selection of participating schools into M@t.abel+ based on unobservables.
Booth Olson et al, 2012	Santa Ana Unified School District	9 middle schools and 6 high schools	72 secondary English teachers	Year 1 : 72 classes (34 treated classrooms and 38 control classrooms) Year 2: 66 classes (31 treated and 35 control)	Although 30% of the teacher sample was lost to attrition from Year 1 to Year 2 (from 103 to 72), there was no systematic difference in attrition rates between conditions and no evidence that teachers who remained in the study were different from those who remained on a baseline measure of teacher self-efficacy.	Teacher and class (teachers were clusters randomly assigned to the Pathway Project, and students were randomly assigned to classrooms. Our study was designed to improve power by creating school by grade randomization blocks and then randomly assigning classroom teachers within each block to experimental conditions)	Groups equivalent at the baseline	Only one district. The district has a long history of providing PD: controls could have previous learnings.

Tabella 3 - Validità

Study	N. of district	N. of schools	N. of teachers	N. of students	Attrition	Randomiz Level	Internal validity	External validity
Bos et al., 2012	8 districts	56 schools (26 controls and 26 treated)	338 treated and 303 controls	The study began with test scores for 6.382 grade 6 intervention students and 6.000 grade 6 control students in Year 1. In Year 2, the sample included test scores for 9.230 grade 7 intervention students and 8.272 control students. In Year 3, the sample included test scores for 9.555 grade 8 intervention students and 8.625 grade 8 control students. These data included students who moved into the schools after the study started.	41% treated and 38% control	School level (using an intent-to-treat model). The researchers designed the study to follow two cohorts of students. The first cohort included students whose teachers had access to QTEL summer professional development and coaching in each year the students were in middle school. The second cohort enrolled in middle school a year later and encountered the same teachers the year following the teachers' first access to QTEL (when its impact on teacher instruction might have either matured or worn off).	1) Students and teachers left the schools between random assignment and when outcome data were collected. 2) Teachers in the control group continued to have access to their regular PD activities. 3) school (2 intervention and 1 control) were consolidated during the study period. To maintain the integrity of random assignment, 1/3 of the teachers and students in the consolidated school were randomly selected and considered control group members in the outcome analyses, even though they were treated as intervention group members by QTEL after the consolidation took place: any estimated QTEL impact on this school would likely be attenuated (biased toward zero), causing a small bias in the overall impact estimates across the entire study sample. 4) Classroom observations were conducted using a convenience sample. 5) Nonresponse on teacher surveys might have caused nonresponse bias in estimates based on data from those surveys; teachers who responded in intervention schools may have systematically differed from teachers who responded in control schools.	1. The sample of school districts and schools was not a random sample. 2. Schools and districts participating in the study volunteered for a study in which schools were randomized to receive QTEL or not. This means that the results may not be representative of schools that are fully committed to the QTEL intervention. 3. Participation in QTEL services was not universal. 4. The fidelity of implementation was limited. The delivery of some QTEL services was compromised by school staffing and logistical issues, and tracking of service receipt by QTEL staff was limited. The findings do not generalize to a setting with complete implementation fidelity. 5. While the classroom observation instruments demonstrated acceptable levels of internal consistency and interrater reliability, there is insufficient data available to establish external validity of either the SIOP or the PACO instrument.
Carpenter et al., 1989			40 teachers (20 treated, 20 controls). Controls participated in two two-hour workshops during the instructional year			Teacher and student (Teacher randomized, 12 students - 6 girls and 6 boys - randomly selected from each class)		

Tabella 3 - Validità

Study	N. of district	N. of schools	N. of teachers	N. of students	Attrition	Randomiz Level	Internal validity	External validity
Cavalluzzo et al., 2012	Kentucky	A volunteer sample of 47 schools (30 of which are in rural areas) was randomly assigned to the treatment and control conditions. The study enrolled 25 schools in the 2007/08 school-year (13 treatment and 12 control) and 22 schools in 2008/09 (11 and 11). Eligibility: low-performing schools	63 treated and 60 controls	The intent-to-treat sample consisted of all grade 9 students in treatment and control schools enrolled on September 1 in a course culminating in the completion of algebra I credit. This sample comprised 6,908 students, 61.4 percent of whom attended rural schools. (3396 treated and 3513 controls)	No attrition at the school level	School-level randomization was chosen, as opposed to within-school randomization, because it minimizes the potential for spillover effects from the treatment to control classrooms, which could lead to underestimating the impacts of the hybrid program.	There are limitations due to the fidelity of implementation as 4 of the 24 treatment schools were noncompliant, which means they did not participate in the intervention. Less than 50 percent of the treatment sample (47.6 percent of teachers in the school-year) had high or moderate attendance for the professional development components.	One limitation is that the sample schools volunteered to participate. Other limitations pertain to the students. Data were missing on the PLAN outcome for students who were not promoted, left the Kentucky public school system, or missed testing in grade 10. The response rate for the PLAN was 86 percent for the control group and 84 percent for the treatment group. Finally, the results of the study apply only to grade 9 students enrolled in courses leading to algebra I credit. The results cannot be generalized to algebra I courses that do not fulfill the algebra I credit requirement, such as algebra I part A (the first course in a two-year sequence) or algebra I lab (an elective course).
Cole, 1992	Mississippi	1 intermediate school	12 teachers (6 treated and 6 controls) and their intact classes	268		Teacher		
Duffy et al., 1986	A large urban district		22 teachers (11 treated and 11 controls) and their intact classes. Controls attended a presentation on effective classroom management.	259 (130 treated and 129 controls) in 22 classes		Teacher and classroom (while identified students were low-achieving readers in each class)		

Tabella 3 - Validità

Study	N. of district	N. of schools	N. of teachers	N. of students	Attrition	Randomiz Level	Internal validity	External validity
Garet et al., 2011	6	39	90 41 treatment 49 control	2.132 1.083 treatment 1.049 controls	Not reported (effects computed on ITT)	School (all 7th grade school teacher included)	RCT (do not provide clear computation of attrition)	By design: school with at least 1/3 of students eligible for free or reduced price lunch
Garet et al., 2008	6	90 (30 treat. A 30 treat. B 30 control)	270 (93 treat. A 88 treat. B 89 control)	5.530 (1.983 treat. A 1.738 treat. B 1.809 control)	Not reported (effects computed on ITT)	Teacher	RCT with low attrition level (do not provide computation)	Schools with higher number of childrens eligible for free or reduced price lunch than average urban or urban fringe schools
Gersten et al., 2010	3 (in 3 States)	19 schools	84 40 treatment 44 control	575 273 treatment 302 control	Overall 18.6% Differential 3.6%	School: all teachers in treatment schools included; 7 student randomized for each teacher	RCT Study underpowered to detect effects on students	Generalizability limited to instructional content focus. Another limitation is that, given the complexity of the TSG and variation in its implementation at each school site, is not clear how this variation could have affected the teacher instruction and outcomes.
Glazerman et al., 2010	17 school districts in 13 US states. districts had to meet criteria for size, poverty, and need for induction	418 elementary schools.	1009 teachers. Within each study school, all eligible teachers participated if they were new to the profession, taught in grades K-6, and were not already receiving induction support from a teacher preparation or certification program.		Not reported	School level (for each district) (1) a treatment group whose beginning teachers were offered comprehensive teacher induction or (2) a control group whose beginning teachers received the district's usual, less comprehensive or intensive induction services.	In 10 of the 17 districts, the services were offered to treatment schools for one year only ("one-year districts"). In the remaining 7 districts, services were offered to treatment schools for two years ("two-year districts"). The two sets of districts implemented different versions of the treatment and they were not randomly chosen to implement one or two years of comprehensive induction.	
Greenleaf et al., 2011	California	83 schools (43 treatment 40 control)	105 56 treatment 49 control	Not reported	Not reported	Teacher	RCT	Based on voluntary teacher.

Tabella 3 - Validità

Study	N. of district	N. of schools	N. of teachers	N. of students	Attrition	Randomiz Level	Internal validity	External validity
Grigg 2012	LAUSD (Los Angeles Unified School District): eight district	In the beginning 80 schools were selected; the final sample was 73 (36 experimental group and 37 control group)			9%	The schools in each local district were randomly assigned to the experimental conditions		Very limited generalizability
Hellerj_2012	55 districts (California and Arizona)	137	181 (90 intervention group and 91 control group).	Estimated number of students for 181 teachers is 10.797 students; total students with ATLAST post test were 5.130	23% intervention group teachers and 30% control group teachers	Teacher	Limitations: high sample attrition; the study did not include analyses of classroom implementation of course-related practices;	The findings are based on volunteer teachers and students whose parents provided consent
Kim et al., 2011	Orange County California	15 secondary schools 9 middle schools 6 high schools	103 52 treatment 51 control	2000 English Learner Students (95% Latino)	Not reported	Student (teacher randomized to treatment and control and students randomized to treatment classroom)	RCT (intervention at the classroom level). Groups equivalent at the baseline	Not enough information to compare the schools to nationwide schools (or to a particular subset).
Lee et al., 2008	A large urban district in the southeast US (not specified)	17 elementary schools (9 treatment 8 controls)	42 treated	1.027 treated 966 control	28%	Classroom	RCT (intervention at the classroom level). Groups equivalent at the baseline	All the school were volunteer. Thus limited generalizability. In each school all the teacher has to participate in the PD.
Marek & Methven, 1991			16 elementary teachers treated and 11 controls identified by treated by peer nomination.	270 (10 students from each of the 27 teachers' classrooms)		Student		
Masters et al., 2010	Not specified (from 8 states)	Not specified	255 148 treatment 107 control	Not reported	40%	Teacher	RCT with high attrition rate (40%) in both conditions	All the teachers were volunteer (so not comparable to nationwide teacher). Only few weeks between the completion of the last workshop and the post-survey.

Tabella 3 - Validità

Study	N. of district	N. of schools	N. of teachers	N. of students	Attrition	Randomiz Level	Internal validity	External validity
Matsamura et al., 2010	1 medium-sized urban district in Texas	29 elementary schools 15 treatment 14 control	73 'replacement' 2nd year only teachers 42 treatment 31 controls	1.269	No attrition in the 73 '2nd year only' teachers	Coaches and teachers	RCT with no attrition, so high internal validity	Very limited generalizability: realized in only one district, targeted to a specific subset of school (low-income) and to a particular subset of teacher (the 'replacement').
Matsamura et al., 2012	urban district in the Southwestern United States	26 school (15 intervention and 11 comparison)	At baseline 192 teachers agreed to participate. At the end of the second year 59 CFC teachers and 32 teachers comparison were still participating.	1.706	At the end of the second year more of the 50%	Schools	Attrition; differences in interventions as planned and implemented; lack observational data on coaching quality	Very limited generalizability: realized in only one district, targeted to a specific subset of school (low-income)
McCutchen et al., 2002			44 teachers/classes (23 treatment and 20 comparison)	779: 492 kindergarteners (268 treated and 224 controls) and 287 first graders (157 treated and 130 controls)				Based on voluntary teacher who responded to an invitation.
McGill-Franzen et al., 1999		six schools (kindergartens)	18 teachers (3 for each school). Three groups: training and books (treated), no training and books, no training and no books.	317 students (164 treated, 153 controls)		Teacher	Comparison between training-and-books and no-training-and-no-books teachers	

Tabella 3 - Validità

Study	N. of district	N. of schools	N. of teachers	N. of students	Attrition	Randomiz Level	Internal validity	External validity
Pianta 2008	one state, 24 different school districts		113 pre-kindergarten teacher participated, 61 in consultation group and 52 in the web only group		35%	District	Limitations: randomization occurred at the district level while the results drawn inferences related to effects at the teacher/classroom level. The fairly large number of predictors entered into each model and the repeated analyses on separate dimensions of teacher-child interaction, which resulted in somewhat diminished power and a slight increase in the possibility that the reported results were obtained by chance, although significant effects were found for 3 of the 10. Another limitation concerns the possible selection bias of children whose parents provided consent to allow their children to participate in this study. CLASS dimensions.	Very limited generalizability: realized in only one state
Randel 2011	32 district in Colorado	67 (33 schools in the intervention group and 34 schools in the comparison group)		2.860 students with CASL treatment and 3.379 students without CASL (6239)	35% of the students	Districts, schools	RCT	Very limited generalizability: realized in only one state, schools within each district were recruited, so they are volunteers.
Sailors & Price, 2010	3 districts in Texas	13 schools	44 teachers (17 treated, 27 controls)	527 (200 treated, 327 controls)	10%	School	They used a Difference in difference, so the causal statements are weaker than in a RCT	The generalizability is very limited, because the study included only 44 teachers in 3 districts with a high portion of low-income students.
Santagata 2011	1	5	64 (34 treatment condition and 30 to the control condition.	3.900	20%	Teacher	Attrition	Very limited generalizability

Tabella 3 - Validità

Study	N. of district	N. of schools	N. of teachers	N. of students	Attrition	Randomiz Level	Internal validity	External validity
Saxe et al., 2001	Los Angeles		23 teachers. 3 comparison groups based on a prescreening questionnaire: a) The Integrated Mathematics Assessment (9); b) the Collegial Support (8); c) Traditional instruction (6)				Comparison between Integrated Mathematics Assessment and Traditional Instruction groups	Based on voluntary teacher who responded to an invitation.
Sloan, 1993		seven Midwestern Schools	10 4th and 5th grade teachers	173 students			pre-post test	
Tienken, 2003		1 school in New Jersey	five 4th grade teachers: trated (2); controls (3)	98 students		Teacher		only post test design, students fom only 5 classes no comparable in their baseline
Vaughn et al., 2011	3 districts in Texas and Colorado	6 middle schools	61 classes (34 trated, 27 controls)	782, (400 treated, 382 controls)	No attrition	Student and classroom	Students random assigned to classrooms, classrooms random assigned to treatments. High internal validity.	The generalizability is very limited, because the study included only 6 schools (17 teachers) in 3 districts.