

La valutazione di impatto dei progetti di formazione all'insegnamento dei docenti universitari: quali indicazioni dalle rassegne delle ricerche empiriche?

Impact evaluation of instructional development programmes in higher education. Evidence from empirical research reviews

MARIA LUCIA GIOVANNINI – ALESSANDRA ROSA*

In questo contributo viene affrontata la problematica della valutazione dell'impatto dei progetti di formazione all'insegnamento universitario, una formazione in progressivo aumento in molti Paesi ma diversificata, quantomeno, per contenuti, durata, destinatari e tipo di obbligatorietà. Lo scopo è quello di delineare il concetto di impatto, descrivere i modelli per la sua valutazione e mettere in luce, in un'ottica di analisi critica, impostazioni e criteri usati nelle diverse rassegne delle ricerche empiriche su tale tematica.

This contribution focuses on impact evaluation of teaching development programmes for university teachers. Such programmes while increasingly implemented in many countries, vary considerably in terms of content, duration and target group as well as the extent to which their participation is recognised as a formal requirement. The scope of the article is to highlight the concept of impact, to analyze the impact evaluation models and approaches and to examine critically the criteria used in different reviews of empirical research on this issue.

Parole chiave: formazione all'insegnamento universitario, valutazione di impatto, efficacia, rassegne delle ricerche empiriche

Key words: higher education, instructional development programme, impact evaluation, effectiveness, research reviews

* Il presente contributo è frutto di un lavoro comune delle autrici, tuttavia per quanto riguarda le singole attribuzioni esse risultano così divise: Maria Lucia Giovannini paragrafi 2, 3, 5; Alessandra Rosa paragrafi 1 e 4.

1. Introduzione

Nel contesto dei rapidi e profondi cambiamenti che negli ultimi decenni hanno attraversato la società in cui viviamo, i sistemi di istruzione superiore di tutti i Paesi avanzati sono stati coinvolti da rilevanti processi di trasformazione e dalle sfide connesse alla necessità di farvi fronte, salvaguardando e rafforzando il ruolo svolto dall'università nella promozione dello sviluppo sociale, economico e culturale. Uno scenario complesso, alla luce del quale va inquadrata la professione del docente universitario sempre più articolata tra ricerca, didattica e governance degli Atenei (Galliani, 2011) e va compresa l'esigenza, sempre più avvertita e riconosciuta a livello internazionale, di innovare e diversificare la didattica universitaria attraverso la predisposizione di contesti formativi partecipativi, coinvolgenti, flessibili e maggiormente centrati sul soggetto che apprende. Il miglioramento della qualità dell'insegnamento e degli apprendimenti degli studenti costituisce, infatti, un obiettivo fondamentale del processo di trasformazione delle istituzioni universitarie e complementare a quello, altrettanto prioritario, rappresentato dall'ampliamento dell'accesso all'istruzione superiore.

Il riconoscimento della necessità di supportare istituzioni e docenti nel perseguimento di tali obiettivi ha contribuito ad accrescere l'attenzione riservata al tema della formazione iniziale e continua all'insegnamento universitario. In particolare a partire dagli anni Novanta, in vari Paesi europei ed extra-europei si assiste non solo all'intensificarsi degli studi e delle occasioni di confronto e dibattito scientifico al riguardo, ma soprattutto alla creazione di apposite strutture e al moltiplicarsi di iniziative ed esperienze formative specificamente finalizzate a promuovere lo sviluppo e il potenziamento delle competenze pedagogico-didattiche dei docenti universitari (Kalman, 2008; De Ketele, 2010; Giovannini, 2010, 2011).

La quantità e la varietà delle esperienze condotte a livello internazionale, unitamente all'imporsi crescente dell'obbligo di assicurazione e rendicontazione della qualità in un contesto di risorse sempre più limitate e di politiche gestionali sempre più improntate ai concetti di efficacia ed efficienza, hanno a loro volta contribuito a determinare l'esigenza di valutare l'impatto di tali progetti di formazione e sviluppo professionale, esplorandone l'effettiva utilità e rilevanza ai fini del miglioramento della qualità della didattica e dei processi di insegnamento e apprendimento. A questo proposito va sottolineato che l'insufficienza di valide e affidabili evidenze empiriche circa l'efficacia degli interventi messi in atto costituisce un tema ricorrente nella letteratura sull'argomento. In essa è frequente il rilievo espresso a proposito della sproporzione tra la numerosità delle iniziative attivate, soprattutto in alcuni Paesi come quelli anglosassoni o nordeuropei, e la scarsità di disegni valutativi in grado di coglierne gli effetti nella pratica didattica e di mettere nel contempo in luce gli elementi su cui far leva nella progettazione e realizzazione di nuove iniziative formative.

2. Obiettivi e interrogativi di ricerca

A partire da tali premesse e in connessione alla necessità di accrescere, nel contesto italiano, la conoscenza della tematica qui affrontata, il presente contributo si propone di fare il punto sulle ricerche empiriche che hanno valutato l'impatto dei progetti di formazione all'insegnamento universitario nei diversi Paesi. Al fine di descrivere e mettere in luce impostazioni, modelli, criteri, evidenze e problematiche, si sono considerati non tanto i singoli studi empirici sul tema quanto le relative rassegne, data la loro disponibilità e la crescente importanza ad esse assegnata.

I primi sforzi di ricognizione e di sintesi risalgono all'inizio degli anni Ottanta e si foca-

lizzano su esperienze e ricerche svolte a partire dalla seconda metà degli anni Sessanta. Gli studi presi in esame da Carrol (1980) riguardano progetti formativi specificamente rivolti ai *teaching assistants*, così come quelli degli anni Ottanta analizzati nella rassegna di Abbott, Wulff e Szego (1989), mentre gli studi empirici considerati da Levinson-Rose e Menges (1981) sono relativi a progetti per docenti universitari con vario grado di esperienza. La rassegna di Weimer e Lenze del 1997 si rifà in gran parte ai criteri adottati da Levinson-Rose e Menges, aggiornando però la letteratura mediante l'inclusione di studi condotti negli anni Ottanta e in alcuni anni Novanta. Nel primo decennio del nuovo millennio si registrano inoltre ben cinque rassegne della letteratura sul tema. Una, pubblicata in lingua fiamminga, è di McAlpine (2003); un'altra è quella di Prebble, Hargraves, Leach, Naidoo, Suddaby e Zepke (2004) che prende in esame studi relativi al periodo 1990-2004 e, per il quinquennio successivo, è aggiornata dalla rassegna di Southwell e Morgan (2009); vi è poi la rassegna di Steinert, Mann, Centeno, Dolmans, Spencer, Gelula e Prideaux (2006) relativa al periodo 1980-2002 focalizzata però esclusivamente su progetti formativi rivolti a docenti universitari di area medica; infine, c'è la rassegna realizzata da Stes, Min-Leliveld, Gijbels e Van Petegem (2010) nella quale la ricerca della letteratura non è stata circoscritta a particolari periodi di tempo¹.

Una testimonianza della sempre maggiore importanza attribuita alle rassegne, connessa alla crescente esigenza di effettuare decisioni e azioni basate su evidenze empiriche, è per esempio rintracciabile nell'attenzione che il *Department of Education* degli Stati Uniti ha rivolto negli ultimi anni alla progressiva definizione di standard di qualità specificamente applicabili alle rassegne incentrate sulla valutazione dei programmi educativi. Anche il dibattito sulle evidenze emerse da tali programmi e sul come sintetizzarle è diventato sempre più intenso ed è tuttora assai vivace tra i ricercatori². Connesso al concetto di impatto si pone infatti, tra gli altri, il problema dell'inferenza causale e quello del legame plausibile tra effetti e cause; è questo un problema di cui non possiamo fare a meno di sottolineare l'elevata complessità anche nel caso in cui si tratti di accertare la relazione non solo tra le attività formative rivolte ai docenti e l'apprendimento dei loro studenti, ma anche tra gli interventi dei formatori nei corsi e le performance effettiva dei docenti universitari.

In relazione all'obiettivo di tipo ricognitivo e descrittivo qui privilegiato, e considerando il ruolo che sintesi e rassegne possono giocare nel supportare processi decisionali e interventi operativi fornendo agli *stakeholders* coinvolti la possibilità di fondare su evidenze empiriche le proprie scelte e azioni, gli interrogativi cui ci siamo proposte di dare risposta sono stati i seguenti: Quali le dimensioni e gli aspetti chiamati in causa dal concetto di impatto nell'ambito della formazione dei docenti? Quali i modelli e gli approcci valutativi utilizzati? Quanto sono stati considerati gli eventuali cambiamenti a livello di pratica didattica dei docenti "formati" e che cosa è emerso? Quanto sono state considerate le ricadute sugli apprendimenti degli studenti e, più in generale, sul contesto lavorativo? È possibile ricavare se, quando e perché funzionano le diverse iniziative formative?

- 1 Benché gli studi inclusi nelle rassegne siano tutti concernenti la valutazione dell'impatto dei progetti di formazione all'insegnamento universitario, è assai limitata la loro sovrapposizione. Questo dipende da vari fattori, quali l'arco di tempo considerato, la tipologia di utenti "formati" (per es. gli assistenti nelle rassegne di Carroll e di Abbott et al., i docenti di area medica in quella di Steinert et al.), il tipo di attività formativa (livelli diversi di formalizzazione), le modalità di ricerca e selezione della letteratura.
- 2 Si veda, per esempio, nella rivista *Educational Researcher* l'interessante dibattito sollecitato dal contributo di Slavin (2008) dal titolo *What Works? Issues in Synthesizing Educational Program Evaluations*.

3. Il concetto di impatto e la sua valutazione

Il concetto di impatto, spesso usato nella letteratura sulla formazione come sinonimo di altri termini che però non sono del tutto tali, richiama: a) la dimensione temporale, in quanto effetto/conseguenza che accade a medio/lungo termine o, secondo alcuni, anche a conclusione di un intervento formativo, b) il soggetto/livello coinvolto dall'effetto/conseguenza (individuale e/o organizzativo), c) il legame tra effetto/conseguenza e intervento (diretto e/o indiretto). Dalla considerazione di tali criteri, cui possiamo aggiungere quelli del valore dell'effetto/conseguenza (positivo/negativo) o dell'intenzionalità dell'effetto/conseguenza (previsto/imprevisto), si ricava in modo esplicito che la valutazione dell'impatto non può essere identificata soltanto con l'accertamento delle percezioni dei docenti partecipanti circa l'efficacia dei progetti formativi seguiti o delle loro autovalutazioni circa gli apprendimenti realizzati. Per cogliere effetti e ricadute della formazione nella loro molteplicità e articolazione è necessario andare oltre la superficie rappresentata dai punti di vista dei partecipanti³ e chiedersi quali siano i cambiamenti effettivamente innescati in seguito all'intervento formativo non solo nei docenti stessi, ma anche nei contesti ai quali appartengono e negli altri attori che in essi agiscono e interagiscono. Sono aspetti da cui traspare la complessità della valutazione dell'impatto; non si può infatti trascurare lo scoglio derivante dalla difficoltà, se non addirittura impossibilità, di misurare certi effetti, né sottovalutare il problema di isolare l'impatto effettivamente dovuto all'azione formativa. A ciò va aggiunto, come si è evidenziato, l'uso intercambiabile di termini non sinonimici (sottolineato per es. da Gray e Radloff, 2008), che facilitano gli equivoci e contribuiscono ad aumentare la complessità del tema.

Con l'intento di migliorare e irrobustire la valutazione degli effetti dei programmi di formazione incentrati sulle competenze di insegnamento dei docenti universitari, sono stati proposti modelli valutativi basati su una concezione di impatto multidimensionale. A riferimento è stato ripreso, articolandolo, il modello gerarchico della valutazione della formazione di Kirkpatrick⁴ (1959, 1976, 1994, 2006). Tale modello riconosce la complessità dell'attività in questione proponendo quattro livelli di analisi caratterizzati da profondità e difficoltà crescenti: ciascun livello successivo al primo, infatti, consente una rilevazione più accurata e completa degli effetti dei progetti formativi, ma al tempo stesso implica procedure di raccolta e analisi dei dati maggiormente dispendiose in termini di tempo e di risorse. Il primo e più semplice livello, definito *reaction*, è costituito dal grado di soddisfazione espresso dai partecipanti in relazione all'utilità dei corsi complessivamente considerati e dal gradimento nei confronti di vari aspetti quali le modalità organizzative adottate, i contenuti trat-

- 3 Tali punti di vista possono tuttavia riflettere impostazioni molto diverse. Per esempio, vi sono ricercatori che, ritenendo indispensabile la valutazione dell'impatto della formazione ma assai difficile la sua realizzazione, per inferire l'impatto propongono sì l'uso di un "questionario di soddisfazione" ma lo collocano in una prospettiva sistemica della formazione tenendo conto delle dichiarazioni sulla pertinenza dell'azione formativa, sugli apprendimenti acquisiti e sull'effettiva realizzazione del loro transfer nel contesto professionale (Gerard, 2003). In tal caso i dati rilevati non corrispondono a quello che viene identificato solitamente come primo livello della formazione.
- 4 Kirkpatrick è considerato il padre dell'ambito della valutazione della formazione. Il termine impatto non compare nella denominazione di nessuno dei quattro livelli di esiti della formazione del suo modello; questo, tuttavia, costituisce il framework di molti modelli utilizzati nella valutazione di impatto dei progetti di formazione all'insegnamento dei docenti universitari. Essendoci limitati in questo contributo a far riferimento ai modelli considerati/delineati per tale tipo e contesto di valutazione, non vengono descritte le impostazioni sull'impatto di importanti studiosi della valutazione di progetti, quali per esempio Stufflebeam (1971, 2007) e Scriven (1972, 1991).

tati, le metodologie didattiche utilizzate, la qualità dei formatori e delle relazioni entro il gruppo. Tali aspetti, per quanto inadeguati a cogliere l'effettivo impatto degli interventi formativi, risultano determinanti dal punto di vista della motivazione all'apprendimento dei partecipanti. Con il secondo livello, definito *learning*, si entra nel vivo degli effetti prodotti dai programmi sui destinatari e beneficiari diretti della formazione, in quanto l'oggetto della valutazione risulta costituito dal cambiamento nelle conoscenze, abilità e atteggiamenti dei partecipanti. Il terzo livello, definito *behavior change*, concerne invece l'applicazione delle competenze sviluppate durante la formazione per migliorare il proprio operato, dunque la loro traduzione in comportamenti concreti nella situazione lavorativa (*transfer on the job*). La considerazione di tali aspetti nelle attività di valutazione risulta fondamentale in quanto elevati livelli di soddisfazione e apprendimento non garantiscono, in modo automatico, il cambiamento delle prassi e dei comportamenti messi in atto: sul *transfer* incidono infatti diversi fattori, tra cui il supporto ricevuto nel periodo successivo alla conclusione delle attività formative. Il quarto e ultimo livello di valutazione, definito *results*, prende infine in esame le misure dell'impatto della formazione sulle finalità e sugli obiettivi dell'organizzazione. Per quanto complessi da misurare e valutare, l'importanza di considerare gli aspetti di tale livello si lega in particolare a quanto appena affermato circa i fattori che incidono sull'applicazione degli apprendimenti. Come evidenziano anche Knight e Trowler (2000), in mancanza di un impatto a livello istituzionale, e dunque di supporto e riconoscimento da parte del proprio contesto professionale, i cambiamenti innescati dai progetti formativi nei singoli partecipanti rischiano infatti di essere riassorbiti nelle prassi e nelle dinamiche tradizionali.

Rifacendosi ai livelli di Kirkpatrick, nel loro modello Hoyt e Howard (1978) differenziano tre tipi di dati da utilizzare negli studi valutativi: quelli che descrivono le reazioni dei partecipanti all'esperienza formativa, i dati relativi ai cambiamenti nei comportamenti e quelli che costituiscono una diretta evidenza del miglioramento nell'efficacia didattica. Ancora più evidente è la trasposizione dei livelli di Kirkpatrick nel modello gerarchico di Smith e Beno (1993); vengono infatti considerati i dati relativi a: 1) la reazione dei partecipanti al progetto formativo dopo ogni attività, 2) i cambiamenti delle loro conoscenze relative all'insegnamento e all'apprendimento, 3) le modificazioni dei loro comportamenti e atteggiamenti, 4) l'impatto sugli studenti e sull'organizzazione. Gli effetti su quest'ultima non vengono analizzati in alcuni altri modelli di anni successivi, per esempio quello di Chism e Szabó (1997) che, oltre al grado di soddisfazione dei docenti partecipanti, fa riferimento alle ricadute dell'intervento formativo sui docenti partecipanti e sui loro studenti.

Anche nel modello valutativo di Guskey (2000, 2002) si ritrovano i livelli della *reaction* e del *learning* dei partecipanti; il terzo livello del *behaviour change* si trasforma invece nell'uso di nuove conoscenze e abilità e il quarto viene specificato come outcomes di apprendimento degli studenti e non soltanto come risultati desiderati. Tra il secondo e il terzo livello, inoltre, viene aggiunto quello del cambiamento e supporto a livello dell'organizzazione; la valutazione dei cambiamenti prodotti a seguito degli interventi formativi rivolti ai docenti viene infatti interpretata anche come documentazione del supporto a questi fornito in termini materiali (risorse e incentivi) e simbolici (incoraggiamento, riconoscimento e valorizzazione) prima e durante la partecipazione alla formazione.

Kreber e Brook (2001), dal canto loro, identificano sei livelli di valutazione dell'impatto, dal più basso al più alto: 1) percezione e soddisfazione della formazione da parte dei partecipanti, 2) convinzioni dei partecipanti sull'apprendimento e insegnamento, 3) performance d'insegnamento dei partecipanti, 4) percezioni da parte degli studenti della performance d'insegnamento, 5) apprendimenti degli studenti, 6) cultura dell'istituzione (compreso il cambiamento istituzionale). Si tratta del modello adottato in diversi studi, come pure nelle rassegne di Prebble et al. (2004) e di Southwell e Morgan (2009).

Nella rassegna di Levinson-Rose e Menges (1981) e in quella di Weimer e Lenze (1997) sono invece state utilizzate cinque categorie, di cui tre riferite ai cambiamenti dei docenti partecipanti alla formazione (atteggiamenti, conoscenze, abilità) e due a quelli dei loro studenti (atteggiamenti e apprendimenti). Quanto alle rassegne più recenti, sono state operate in esse alcune rielaborazioni a partire dalla versione parzialmente modificata del modello di Kirkpatrick proposta da Freeth et al. (2003). In particolare sono stati rivisti il secondo e il quarto livello di analisi, cioè quelli definiti *learning* e *results*. Quest'ultimo livello è stato inteso nei termini di "apprendimento organizzativo" e di effetti prodotti sui beneficiari indiretti, vale a dire di ricadute sulla cultura, sul clima e sulle strategie istituzionali, nonché di impatto sugli studenti, aspetto già evidenziato da Kreber e Brook (2001) e da Guskey (2002), isolato dalla più ampia categoria dell'impatto istituzionale e articolato in vari tipi di outcomes relativi a percezioni, approcci e risultati di apprendimento. In relazione al livello costituito dagli apprendimenti dei docenti partecipanti, Stes et al. (2010) hanno recuperato e aggiunto il cambiamento nelle concezioni relative all'insegnamento, in quanto la presa di consapevolezza e lo sviluppo delle teorie implicite dei docenti rientrano spesso tra gli obiettivi formativi dei programmi e, aspetto che noi condividiamo, vengono ritenute di importanza critica nel favorire un'effettiva modificazione delle prassi didattiche.

Gli autori delle rassegne hanno escluso dalle loro sintesi le ricerche empiriche incentrate esclusivamente su dati relativi al livello *reaction*, e dunque alle percezioni dei docenti circa la positività e utilità dell'esperienza e alla loro soddisfazione; tale livello nella rassegna di Stes et al. (2010) è stato addirittura eliminato dal framework adottato. Si tratta infatti di una dimensione che, in base a una posizione condivisa da vari autori (per es. Holton, 1996, nel cui modello vengono considerati i risultati relativi all'apprendimento, alla performance e all'organizzazione), non fornisce propriamente misure di impatto.

Nei framework per la valutazione dell'impatto degli interventi formativi rientrano inoltre altri aspetti quali, per esempio, la loro durata, il tipo di obbligatorietà e di incentivi/riconoscimento, il livello di formalizzazione e le metodologie adottate, gli anni di esperienza e altre caratteristiche dei partecipanti. Inoltre, benché lo scopo dei progetti di formazione all'insegnamento sia quello di migliorare la qualità dell'apprendimento degli studenti accrescendo la qualità dell'insegnamento, in relazione al modello/quadro teorico di sviluppo professionale dei docenti assunto (per es. Gilbert e Gibbs, 1999) il focus del/i cambiamento/i perseguito/i dagli interventi formativi può essere posto su uno o più aspetti relativi all'attività di insegnamento (per es. comportamento didattico, cambiamento concettuale, pratica riflessiva) oppure può essere rivolto anche ad aspetti relativi agli studenti (per es. loro approccio all'apprendimento). Già da queste esemplificazioni si può comprendere la varietà delle misure e degli indicatori di impatto utilizzati nei vari studi empirici analizzati nelle diverse rassegne e la conseguente complessità di una loro comparazione.

4. Impostazioni ed esiti delle rassegne⁵

In base alla nostra ricognizione della letteratura internazionale, la prima rassegna specificamente incentrata sugli effetti dei programmi di formazione all'insegnamento per docenti universitari è quella effettuata da Carroll (1980), che si focalizza in particolare su iniziative

5 Rispetto alle rassegne precedentemente citate, non si considerano qui i risultati della rassegna di McAlpine (2003) in quanto pubblicata in lingua fiamminga e relativa soltanto a sette studi.

rivolte ai *teaching assistants* (TAs). I 24 studi valutativi considerati nella rassegna, dalla quale sono state escluse le indagini incentrate esclusivamente su dati relativi alla soddisfazione dei partecipanti, sono stati classificati e analizzati in base ai tipi di outcomes rilevati, suddivisi in *TA variables* (cambiamenti nelle conoscenze, atteggiamenti e comportamenti dei partecipanti) e *student variables* (cambiamenti nelle attitudini e nei risultati degli studenti e nelle loro percezioni sui docenti e sui corsi), e in base al tipo di disegno adottato (*correlational, pre-experimental, quasi-experimental, true-experimental*). Per quanto concerne i risultati emersi, la maggior parte degli studi considerati rivela effetti positivi dei programmi formativi sia sui docenti (in 14 casi su 17) sia, in misura minore (in 11 casi su 19), sui loro studenti, ma l'autore sottolinea che i dati disponibili non consentono di trarre conclusioni definitive in quanto la scarsità e la "debolezza" metodologica degli studi effettuati rendono necessari ulteriori approfondimenti e conferme attraverso l'uso di disegni valutativi più articolati, robusti e rigorosi. È quanto mettono in evidenza anche Abbott *et al.* (1989) nella loro rassegna relativa a 13 studi, tutti degli anni Ottanta, concernenti i *teaching assistants*.

Quasi contemporanea alla rassegna di Carroll è quella effettuata da Levinson-Rose e Menges (1981), limitata al quindicennio 1965-1980 e in questo caso focalizzata sull'impatto di iniziative rivolte non a uno specifico *target group*, bensì a docenti universitari di vario grado ed esperienza. Rispetto alla precedente *review* cambiano anche, almeno in parte, i criteri adottati per l'individuazione, la classificazione e l'analisi dei 71 studi presi in esame. Essi si basano da un lato sulla tipologia di intervento di *instructional improvement* oggetto di valutazione (sovvenzioni per progetti di innovazione della didattica, workshops e seminari, analisi delle valutazioni espresse dagli studenti, esercitazioni pratiche, uso di film e videotape) e, dall'altro lato, su 5 categorie relative al tipo di effetto rilevato (cambiamenti negli atteggiamenti, conoscenze e prassi dei docenti e cambiamenti negli atteggiamenti e negli apprendimenti degli studenti). Le indagini incentrate esclusivamente su dati relativi a una o a entrambe le prime due categorie di outcome, cioè agli atteggiamenti e alle conoscenze dei partecipanti, sono state escluse dalla rassegna. Nel complesso, i risultati degli studi analizzati indicano effetti positivi per il 78% degli interventi presi in esame, anche se tale percentuale si abbassa al 62% quando si considerano esclusivamente quelli reputati dagli autori come maggiormente attendibili; sebbene nella maggior parte dei casi emergano ricadute positive, molte delle ricerche svolte vengono infatti giudicate di bassa qualità. In particolare, nella rassegna vengono messe in rilievo le seguenti criticità: gran parte dei dati si colloca al livello più superficiale rappresentato dalla soddisfazione e dalle percezioni dei partecipanti, mentre sono necessarie misure di performance volte a rilevare in modo più diretto e rigoroso i cambiamenti effettivi nei comportamenti e negli approcci adottati; non si tiene conto delle differenze individuali tra docenti e tra studenti; le variabili dipendenti considerate sono definite e operazionalizzate in maniera eterogenea e dunque poco comparabile; mancano dati di natura qualitativa; gli studi svolti sono spesso frammentati e isolati e mancano progetti di ricerca di tipo interistituzionale.

La rassegna effettuata da Weimer e Lenze (1997), che rispetto alle altre si caratterizza per un approccio di tipo più qualitativo e descrittivo, si propone come aggiornamento di quella precedente. I criteri generali su cui basa la rassegna sono gli stessi adottati da Levinson-Rose e Menges, cioè il tipo di intervento e di effetti oggetto di valutazione, mentre sono almeno in parte diverse le specifiche categorie di analisi considerate; il cambiamento nel tempo delle tipologie di attività e iniziative rivolte al miglioramento dell'insegnamento universitario ha spinto infatti gli autori a rivedere parzialmente le categorie relative a tale criterio (workshops, seminari e programmi; consulenza individuale; sovvenzioni per progetti di innovazione della didattica; distribuzione di materiali; supporto tra colleghi). Per quanto concerne invece il secondo criterio, le categorie restano le stesse, ma in questo caso non vengono escluse le

indagini incentrate esclusivamente su dati relativi alle prime due (atteggiamenti e conoscenze dei docenti). I risultati degli studi valutativi presi in esame sono ancora una volta complessivamente positivi, ma si sottolinea che le evidenze disponibili non consentono di trarre conclusioni definitive e generalizzabili circa l'efficacia degli interventi. Nonostante gli anni di distanza dalla *review* di Levinson-Rose e Menges, sembrano infatti permanere le stesse problematiche e criticità relative agli orientamenti e all'impianto delle ricerche, come il predominante ricorso a percezioni e punti di vista dei docenti e la carenza di evidenze empiriche più robuste e attendibili circa il cambiamento nelle loro pratiche di insegnamento e nei risultati degli studenti. In conclusione gli autori evidenziano dunque la necessità di intensificare la ricerca sul campo, di adottare disegni più sofisticati, di utilizzare misure più dirette e rigorose e di ampliare le metodologie di indagine superando un approccio prettamente quantitativo.

La successiva rassegna è quella effettuata da Prebble et al. (2004) su richiesta del Ministro dell'Educazione della Nuova Zelanda, la quale prende in esame la letteratura pubblicata nel periodo 1990-2004. I criteri generali adottati per la classificazione e l'analisi dei 33 studi individuati sono, come nei due casi precedenti, il tipo di iniziativa e di effetti oggetto di valutazione, mentre sono parzialmente diverse le specifiche categorie di analisi considerate. Per quanto riguarda il primo criterio, le categorie identificate da Levinson-Rose e Menges e da Weimer e Lenze vengono aggiornate alla luce dei nuovi sviluppi nel campo (corsi brevi; programmi estesi nel tempo; formazione *in situ*; consulenza, valutazione tra pari e mentoring; analisi delle valutazioni espresse dagli studenti); per quanto riguarda invece il secondo criterio gli autori, facendo esplicito riferimento al modello di Kreber e Brook (2001), aggiungono alle categorie di impatto precedentemente considerate un'ulteriore dimensione relativa agli effetti a livello istituzionale, cioè sulla cultura dell'organizzazione di appartenenza dei docenti partecipanti. Vengono inoltre incluse nella rassegna le indagini riconducibili alle seguenti tipologie: studi quantitativi condotti su ampi campioni e su scala interistituzionale; studi quantitativi ma condotti all'interno di una singola istituzione; studi qualitativi sulla base di dati provenienti da più istituzioni; studi qualitativi riguardanti singole istituzioni; studi di natura teorica ma che sintetizzano gli esiti di più studi quantitativi. I risultati della rassegna, anche in questo caso complessivamente positivi circa l'impatto delle iniziative considerate, portano gli autori a concludere che attraverso una molteplicità di interventi di formazione e sviluppo professionale i docenti possono essere supportati nel miglioramento del proprio insegnamento. Si sottolinea, tuttavia, che i corsi brevi tendono ad avere un impatto limitato sul cambiamento delle concezioni e delle prassi dei docenti e risultano più efficaci per trasmettere informazioni o per la formazione in specifiche abilità e tecniche, mentre le altre tipologie di intervento sembrano in grado di produrre più rilevanti effetti sulla qualità della didattica e degli apprendimenti degli studenti. I risultati emersi, inoltre, vengono interpretati con cautela, in quanto non solo le varie ricerche condotte risultano difficilmente comparabili, ma in gran parte si limitano a valutare l'impatto dei programmi basandosi su opinioni e autovalutazioni dei partecipanti senza esaminare in modo più rigoroso e sistematico i cambiamenti nelle loro prassi e nei risultati degli studenti. Tali conclusioni vengono confermate e messe in evidenza dalla rassegna di Southwell e Morgan (2009), che prende in esame 28 studi aggiornando per il periodo 2004-2009 la rassegna di Prebble et al. (2004) e adottandone impostazione e finalità.

Steinert et al. (2006) hanno effettuato una rassegna che si differenzia dalle altre in quanto incentrata su programmi di formazione all'insegnamento specificamente rivolti a docenti universitari di area medica. Tale rassegna, che ha incluso 53 studi individuati mediante il ricorso a tre diverse banche dati di natura disciplinare e circoscrivendo la ricerca al periodo 1980-2002, è stata guidata dal *conceptual framework* sviluppato dai ricercatori che si caratterizza

per i seguenti aspetti: riconosce che l'insegnamento fa parte di un più ampio insieme di attività e ruoli assegnati agli accademici; sottolinea che l'impatto degli interventi volti a migliorarne l'efficacia risulta mediato da numerosi fattori di natura contestuale; evidenzia che tale impatto può e deve essere osservato a vari livelli. Per quanto concerne in particolare quest'ultimo aspetto, gli effetti dei programmi vengono classificati e analizzati in base a una versione leggermente riadattata del modello di Kirkpatrick (1998) così come modificato da Freeth et al. (2003), vale a dire considerando le seguenti categorie: *reaction* (punti di vista dei docenti partecipanti sull'esperienza formativa, organizzazione, contenuti, metodologia adottata); *learning* (distinguendo i cambiamenti nei loro atteggiamenti da quelli nelle conoscenze e abilità); *behavior* (cambiamenti nelle loro prassi e comportamenti); *results* (impatto sull'organizzazione e impatto sui colleghi e sugli studenti). Gli studi che riportano esclusivamente dati relativi alla prima dimensione sono stati esclusi dalla rassegna. Per quanto riguarda invece il tipo di intervento, a differenza delle altre *review*, gli autori hanno scelto di focalizzarsi principalmente sulle modalità più formali, collettive e strutturate di sviluppo professionale considerando le seguenti categorie di analisi: workshops; corsi brevi; cicli di seminari; programmi estesi nel tempo. Nel complesso, dalla rassegna emerge non solo un elevato grado di soddisfazione per tali iniziative, ma soprattutto che esse sembrano in grado di promuovere il miglioramento dei docenti e della didattica (in particolare quelle di durata più estesa); per quanto concerne l'impatto sugli studenti e quello organizzativo gli autori affermano che gli studi svolti sono ancora troppo pochi, ma anche in questo caso si riscontrano effetti complessivamente positivi. Ancora una volta, tuttavia, si sottolineano i limiti degli studi condotti e si evidenziano esigenze molto simili a quelle messe in luce dalle rassegne precedenti, quali ad esempio: utilizzare disegni più robusti e rigorosi; integrare metodi quantitativi e qualitativi; fare maggiore riferimento a misure di tipo *performance-based*; utilizzare varie fonti di dati e vari strumenti; aumentare la comparabilità tra gli studi svolti, ad esempio impiegando strumenti standardizzati; promuovere progetti di ricerca di tipo collaborativo e interistituzionale; indagare l'impatto a lungo termine dei programmi; dedicare più attenzione all'impatto istituzionale e a quello sull'apprendimento degli studenti.

L'ultima rassegna comparsa in ordine di tempo è quella di Stes et al. (2010), che rispetto alla precedente torna ad occuparsi di iniziative di *instructional development* di vario tipo e rivolte a docenti universitari di diverse discipline e vari anni di esperienza. Il *framework* alla base di tale rassegna, che ha incluso 36 studi reperiti attraverso il database ERIC senza limitare temporalmente la ricerca, è costituito anche in questo caso da un adattamento del modello di Kirkpatrick, ma rispetto alla rassegna di Steinert e collaboratori vengono introdotti maggiori aggiustamenti: il primo livello di valutazione (*reaction*) viene eliminato, in quanto si ritiene che la soddisfazione dei partecipanti non dica nulla circa l'impatto degli interventi; al secondo livello (*learning*) viene aggiunta una sub-categoria relativa ai cambiamenti nelle concezioni sull'insegnamento, ritenute di importanza critica ai fini di una modificazione delle prassi dei docenti; gli effetti sugli studenti vengono separati dalla più ampia categoria dell'impatto istituzionale, andando a costituire un livello a sé stante. In conclusione, il modello di riferimento della rassegna risulta articolato nei seguenti livelli e sotto-livelli: 1) cambiamenti nei docenti (atteggiamenti, concezioni, conoscenze, abilità e comportamenti); 2) cambiamenti negli studenti (percezioni sui docenti e sui corsi, approcci allo studio e risultati di apprendimento); 3) impatto istituzionale (cambiamenti organizzativi). Considerando complessivamente gli studi presi in esame e i vari livelli di impatto, dalla rassegna emergono in 86 casi effetti positivi, in 17 casi effetti solo parzialmente positivi e, soltanto in 3 casi, non si riscontra alcun tipo di effetto. Sebbene dunque l'impatto delle iniziative considerate appaia generalmente positivo, anche in questo caso si suggerisce cautela nell'interpretazione dei risultati in ragione di una serie di questioni problematiche, già evidenziate dalle precedenti

rassegne, relative alle caratteristiche della letteratura disponibile. Per quanto riguarda i tipi di *outcomes* misurati, ad esempio, la maggior parte degli studi considerati si concentra sugli apprendimenti (27) e sui comportamenti (31) dei docenti partecipanti, mentre minore attenzione viene dedicata all'impatto sugli studenti (12) e a quello istituzionale (9); in quasi tutti i casi, inoltre, si fa riferimento a dati di tipo *self-report*, cioè a quanto percepito e dichiarato dai partecipanti stessi. Dal punto di vista metodologico si sottolinea l'esigenza di disegni di ricerca più rigorosi e affidabili (esperimenti e quasi-esperimenti, *mixed-method approach*), di studi longitudinali e di una maggiore comparabilità tra le ricerche e tra i risultati prodotti. Viene infine indicata la necessità di indagare e approfondire la questione delle differenze di impatto connesse alle caratteristiche delle varie iniziative formative, quali durata, tipologia e destinatari; rispetto a ciò, i risultati della sintesi effettuata suggeriscono che gli interventi più estesi nel tempo incidono maggiormente sulla modificazione dei comportamenti dei partecipanti e che, rispetto ad altre forme più innovative, gli interventi di stampo tradizionale (corsi di varia durata) producono un impatto più positivo a livello degli studenti, ma si tratta di riscontri che richiedono ulteriori indagini di approfondimento e conferme.

Considerazioni conclusive

Esaminando nel loro insieme caratteristiche ed esiti delle rassegne disponibili circa l'impatto delle iniziative volte a potenziare le competenze di insegnamento dei docenti universitari è possibile effettuare, a nostro avviso, alcune considerazioni generali.

In primo luogo si tratta di *review* che, pur perseguendo obiettivi molto simili, si differenziano per vari aspetti quali le scelte effettuate in relazione alle tipologie di intervento considerate e ai criteri di analisi adottati, rendendo dunque problematica la comparazione dei risultati. Anche la scarsa sovrapposizione esistente tra gli studi e le ricerche che ciascuna di esse ha preso in esame può essere vista come un segnale di tali differenze di approccio. In secondo luogo, pur nella diversità di impostazioni e opzioni, tutte le rassegne prese in esame giungono a conclusioni simili rispetto agli effetti e ricadute degli interventi considerati ma anche rispetto ai limiti, alle perplessità e ai problemi (soprattutto di natura metodologica) che inficiano validità e attendibilità dei risultati emersi. Sembrerebbe insomma che, nel corso del tempo, non si siano fatti sufficienti sforzi in direzione di una valutazione di impatto più sistematica e rigorosa o, forse più realisticamente, non si è riusciti a perseguire tale traguardo. A meno che ci si limiti a rilevare le percezioni di impatto dichiarate dai soggetti coinvolti, vale a dire il cosiddetto impatto percepito, le difficoltà da affrontare e superare sono molteplici e per certi aspetti insormontabili. Per poter valutare l'impatto della formazione in termini di inferenza causale o di legame plausibile tra effetti e cause, occorrerebbe riuscire a comparare ciò che è accaduto con ciò che avrebbe avuto luogo in assenza dell'intervento formativo ma, al di là della indisponibilità di tali dati, resterebbe comunque il problema di come riuscire a precisare in quale misura tale differenza sia determinata dall'intervento della formazione e/o da quello di altre variabili.

Un cambiamento in positivo emerge invece a livello dei modelli e presupposti teorici alla base delle rassegne, che in quelle più recenti appaiono improntati a una concezione più complessa e articolata di impatto; esso infatti arriva a inglobare le ripercussioni sul contesto istituzionale, anche se poi nella pratica tale dimensione risulta ancora scarsamente esplorata.

Alla luce degli obiettivi del presente contributo l'aspetto che tuttavia ci appare più importante consiste nello scarso approfondimento, negli studi considerati dalle rassegne, di analisi longitudinali multimetodo che mirino a rilevare i processi di cambiamento durante le attività formative e a lungo termine, nonché di analisi di approfondimento delle relazioni

esistenti tra le caratteristiche dei percorsi formativi, nonché dei partecipanti, e l'impatto degli interventi. A questo si lega, a nostro avviso, la difficoltà di trarre e ricavare dalle esperienze condotte implicazioni e indicazioni operative per la progettazione e realizzazione di analoghe iniziative. Come affermano anche Stes et al. (2010), gli studi svolti per valutare l'impatto della formazione spesso riportano solo generiche e scarse informazioni sulla natura delle iniziative prese in esame, e ciò limita la possibilità di individuare quali elementi e fattori contribuiscano maggiormente a rendere efficaci gli interventi formativi o siano maggiormente collegati ad essi.

Tanto la necessità di indagare l'impatto a livello istituzionale quanto quella di soffermarsi sulla natura degli interventi si legano all'esigenza, messa in luce da alcuni autori (ad es. Chatterji, 2008), di collocare la valutazione di impatto dei programmi formativi in una prospettiva sistemica, cioè allargata alla considerazione di fattori di influenza esterni ai programmi stessi, e contestualizzata, cioè volta a leggere e interpretare i risultati alla luce delle specificità dei contesti formativi e a comprendere gli aspetti e le modalità che possono aver contribuito a determinarli, in relazione a una teoria del cambiamento. In tale ottica, infatti, gli studi valutativi e le rassegne volte a sintetizzarne gli esiti possono acquisire effettiva utilità e rilevanza nel supportare politiche, decisioni e azioni che, in questo caso, sono relative agli interventi di formazione – iniziale e continua – all'insegnamento universitario.

Riferimenti bibliografici

- Abbott R. D., Wulff D. H., Szego C. K. (1989). Review of research on TA training. In J. D. Nyquist, R. D. Abbott, D. H. Wulff (Eds.), *Teaching assistant training in the 1990s. New directions for teaching and learning* (pp. 111-124). San Francisco: Jossey-Bass.
- Carroll J.G. (1980). Effects of training programs for university teaching assistants. *Journal of Higher Education*, 51(2), 167-183.
- Chatterji M. (2008). Synthesizing evidence from impact evaluations in education to inform action. *Educational Researcher*, 37(1), 23-26.
- Chism N.V.N., Szabó B.S. (1997). How faculty development programs evaluate their services. *Journal of Faculty, Program and Organization Development*, 15(2), 55-62.
- De Ketele J.-M. (2010). La pédagogie universitaire: un courant en plein développement. *Revue Française de Pédagogie*, 3(172), 5-13.
- Freeth D., Hammick M., Koppel I., Reeves S., Barr H. (2002). *A critical review of evaluations of interprofessional education*. London: Learning and Teaching Support Network: Health Sciences and Practice.
- Galliani L. (Ed.) (2011). *Il docente universitario. Una professione tra ricerca, didattica e governance degli Atenei*. Tomo I. Lecce: Pensa MultiMedia.
- Gerard F.-M. (2003). L'évaluation de l'efficacité d'une formation. *Gestion 2000*, 20(3), 13-33.
- Gilbert A., Gibbs, G. (1999). A proposal for an international collaborative research programme to identify the impact of initial training on university teachers. *Research and Development in Higher Education*, 21, 131-143.
- Giovannini M.L. (Ed.) (2010). *Insegnare all'Università. Modelli di formazione in Europa*. Bologna: CLUEB.
- Giovannini M.L. (2011). La formazione all'insegnamento dei professori universitari. In L. Galliani (Ed.), *Il docente universitario. Una professione tra ricerca, didattica e governance degli Atenei*. (Tomo I, pp. 63-78). Lecce: Pensa MultiMedia.
- Gray K., Radloff A. (2008). The idea of impact and its implications for academic development work. *International Journal for Academic Development*, 13(2), 97-106.
- Guskey T. R. (2000). *Evaluating professional development*. Thousand Oaks, CA: Corwin Press.
- Guskey T.R. (2002). Does it make a difference? Evaluating professional development. *Educational Leadership*, 59(6), 45-51.
- Holton E. F. (1996). The flawed four-level evaluation model. *Human Resource Development Quarterly*, 7(1), 5-21.

- Hoyt D.P., Howard G. S. (1978). The evaluation of faculty development programmes. *Research in Higher Education*, 8(1), 25-38.
- Kalman A. (2008). *Case studies in the development and qualification of the university teachers in Europe*. Southampton, UK: NETTLE Project Publications.
- Kirkpatrick D. L. (1959). Techniques for evaluating training programs. *Journal of American Society for Training and Development*, 13(11), 3-9.
- Kirkpatrick D. L. (1976). Evaluation of training. In R. L. Craig (Ed.), *Training and Development Handbook: A guide to human resources development*. New York: McGraw-Hill.
- Kirkpatrick D. L. (1994). *Evaluating training programs. The four levels*. San Francisco, CA: Berrett-Koehler.
- Kirkpatrick D. L., Kirkpatrick J. D. (2006). *Evaluating training programs. The four levels* (3th ed.). San Francisco, CA: Berrett-Koehler.
- Knight P.T., Trowler P.R. (2000). Department-level cultures and the improvement of learning and teaching. *Studies in Higher Education*, 25(1), 69-83.
- Kreber C., Brook P. (2001). Impact evaluation of educational development programmes. *International Journal for Academic Development*, 6(2), 96-108.
- Levinson-Rose J., Menges R.J. (1981). Improving college teaching: a critical review of research. *Review of Educational Research*, 51(3), 403-434.
- McAlpine L. (2003). Het belang van onderwijskundige vorming voor studentgecentreerd onderwijs: de praktijk geëvalueerd. In N. Druine, M. Clement, K. Waeytens (Eds.), *Dynamiek in het hoger onderwijs. Uitdagingen voor op onderwijsondersteuning* (pp. 57-71). Leuven: Universitaire Pers.
- Prebble T., Hargraves H., Leach L., Naidoo K., Suddaby G., Zepke N. (2004). *Impact of student support services and academic development programmes on student outcomes in undergraduate tertiary study: a synthesis of the research*. Wellington: New Zealand Ministry of Education.
- Scriven M. (1972). Pros and cons about goal-free evaluation. *Evaluation Comment*, 3(4), 1-7.
- Scriven M. (1991). *Evaluation thesaurus*. Thousand Oaks, CA: Sage.
- Slavin R.E. (2008). What works? Issues in synthesizing educational program evaluations. *Educational Researcher*, 37(1), 5-14.
- Smith C., Beno B. (1993). *Guide to staff development evaluation*. Sacramento, CA: Community College League of California.
- Southwell D., Morgan W. (2009). *Leadership and the impact of academic staff development and leadership development on student learning outcomes in higher education. A review of the literature*. A report for the Australian Learning and Teaching Council (ALTC). Brisbane: Queensland University of Technology.
- Steinert Y., Mann K., Centeno A., Dolmans D., Spencer J., Gelula M., Prideaux D. (2006). A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide n. 8. *Medical Teacher*, 28(6), 497-526.
- Stes A., Min-Leliveld M., Gijbels D., Van Petegem P. (2010). The impact of instructional development in higher education: the state-of-the-art of the research. *Educational Research Review*, 5(1), 25-49.
- Stufflebeam D.L. (1971). The relevance of the CIPP evaluation model for educational accountability. *Journal of Research and Development in education*, 5(1), 19-25.
- Stufflebeam D.L., Skinkfield A. (2007). *Evaluation: theory, models and applications*. San Francisco, CA: Jossey-Bass.
- Weimer M., Lenze L.F. (1997). Instructional interventions: a review of the literature on efforts to improve instruction. In R. Perry, J. Smart (Eds.), *Effective teaching in higher education: research and practice* (pp. 205-240). New York: Agathon.