

## **Il modello dei quattro livelli di Kirkpatrick per valutare la formazione continua: anelli e legami deboli di una “catena di evidenze”.**

### **Kirkpatrick’s four-level model for evaluating training programs: the weak links in a "chain of evidence”.**

Alessandra Rosa, Università degli Studi di Bologna.

#### **ABSTRACT ITALIANO**

Il contributo presenta un’analisi in prospettiva storica e critica del “modello dei quattro livelli” messo a punto da Donald Kirkpatrick, riferimento teorico-metodologico tra i più noti e utilizzati a livello internazionale nel campo della valutazione della formazione continua. In particolare, ripercorrendo gli scritti dell’autore in connessione al dibattito scientifico relativo al modello, vengono delineate alcune rilevanti aree di criticità connesse ai suoi presupposti e al suo impiego, ponendo il focus dell’analisi su questioni controverse come quelle concernenti le relazioni tra i livelli e il *transfer* dell’apprendimento nel contesto professionale. In riferimento alle problematiche discusse, vengono infine proposte alcune riflessioni su possibili implicazioni che se ne possono trarre in termini di prospettive di sviluppo della ricerca e della pratica valutativa nell’ambito della formazione continua.

#### **ENGLISH ABSTRACT**

The paper examines from an historical and critical perspective the "four-level model" developed by Donald Kirkpatrick, a very popular and widely used framework within the field of training programs evaluation. By reviewing the author’s contributions in connection with the scientific debate on the model, an overview of the main critical areas related to its assumptions and use is outlined, focusing the analysis on some controversial issues such as those concerning the relationships between levels and the learning transfer process. Finally, with reference to the problems discussed, some conclusive reflections are proposed on implications that can be drawn in terms of developments in the field of training evaluation research and practice.

## **Introduzione**

Nel quadro di una società e di un’economia basate sulla conoscenza e di un mercato del lavoro complesso e flessibile, la formazione continua – intesa quale formazione specificamente rivolta a soggetti adulti occupati o disoccupati/inoccupati e primariamente finalizzata a promuovere lo sviluppo delle loro competenze professionali (Marcuccio, 2012) – ha acquisito importanza strategica tanto per gli individui quanto per le organizzazioni. Per i primi può costituire una risorsa per far fronte ai cambiamenti richiesti da carriere lavorative sempre meno stabili, lineari e prevedibili, per le seconde rappresenta una fondamentale leva su cui agire per incrementare la propria produttività e

competitività, che sempre più dipendono da investimenti nella qualificazione e crescita del cosiddetto “capitale umano”.

In tale scenario è maturata, negli ultimi decenni, un’attenzione crescente anche per la valutazione dell’efficacia e dell’impatto della formazione continua, necessaria per rilevare informazioni utili a supportare processi decisionali a vari livelli: da quelli più legati alla funzione rendicontativa della valutazione, centrata sui risultati prodotti in rapporto agli obiettivi perseguiti e alle risorse investite, che coinvolgono principalmente decisori politici, enti finanziatori, *management* interno alle organizzazioni; a quelli in cui la valutazione assume una funzione prevalentemente regolativa/migliorativa, centrata sui processi e sui loro legami con i risultati, che chiamano in causa le figure a vario titolo coinvolte nella progettazione ed erogazione della formazione (quali progettisti, coordinatori, formatori, tutor).

L’intenzione di offrire a tale insieme di *stakeholders* «un linguaggio comune e un modo semplice per comunicare cosa significhi valutare e come ciò vada fatto» (Kirkpatrick & Kirkpatrick, 2005, p. 3) (1), a sua volta legata al riconoscimento dell’intrinseca complessità della valutazione nelle sue molteplici dimensioni e funzioni, è ciò che ha mosso Donald Kirkpatrick nel momento in cui – alla fine degli anni Cinquanta del secolo scorso – ha presentato l’approccio articolato in quattro *step* che tuttora costituisce un modello classico e centrale nel campo della valutazione della formazione continua, con particolare riferimento, anche se non esclusivo, all’ambito aziendale. La formazione alle/nelle imprese costituisce infatti il focus principale del modello proposto da Kirkpatrick, nonché l’ambito organizzativo maggiormente rappresentato tra gli esempi e i casi illustrati nei suoi volumi, ma il *framework* è stato utilizzato – con opportuni adattamenti – anche in altri contesti di applicazione (2).

A partire da tali premesse, il contributo propone un’analisi in prospettiva storica e critica del modello volta a mettere in luce, accanto ai suoi innegabili pregi, alcune debolezze e problematiche che rischiano di rimanere sullo sfondo di una facciata di popolarità indiscussa ma vanno invece attentamente considerate nell’ottica di promuoverne un uso consapevole e critico. In primo luogo si ripercorrono le origini del modello, le ragioni della sua rapida affermazione e ci si chiede se e come l’autore, attraverso gli scritti proposti nell’arco di circa cinquant’anni, abbia sviluppato il *framework* inizialmente proposto anche in risposta ai rilievi mossi da altri studiosi. Vengono poi ricostruite alcune rilevanti linee del dibattito sulle aree di criticità connesse ai presupposti del modello e al suo impiego effettivo nelle pratiche valutative correnti, prendendo in esame questioni controverse come quelle inerenti le relazioni tra i livelli e il *transfer* dell’apprendimento nel contesto professionale. In riferimento alle problematiche discusse, vengono infine proposte alcune riflessioni sulle implicazioni derivanti dall’analisi effettuata in termini di prospettive di sviluppo della ricerca e della pratica valutativa nel campo della formazione continua.

## Il modello dei quattro livelli di valutazione della formazione: origini e sviluppi

Con una serie di quattro articoli pubblicati tra il 1959 e il 1960 sulla rivista dell'*American Society for Training and Development*, Donald Kirkpatrick presenta per la prima volta un insieme di tecniche per la valutazione di programmi formativi articolato per *step* successivi e posti in ordine gerarchico destinato a divenire noto, negli anni successivi, come “modello dei quattro livelli”. Ciascun articolo è appunto dedicato a uno dei quattro *step* considerati, mediante i quali l'autore delinea progressivamente una sequenza di attività valutative incentrate su differenti oggetti e caratterizzate da complessità crescente in termini di procedure e risorse necessarie. Il primo livello, denominato *reaction*, si riferisce alle impressioni dei partecipanti sul percorso formativo e mira a valutare il grado in cui l'esperienza – nei suoi vari aspetti relativi alle modalità organizzative, ai contenuti, alla docenza, ai metodi e materiali utilizzati – è stata accolta e vissuta positivamente (Kirkpatrick, 1959a). Il secondo, denominato *learning*, mira a determinare in che misura i partecipanti hanno appreso i principi, le procedure e le tecniche proposti durante il percorso (Kirkpatrick, 1959b), mentre il terzo, denominato *behavior*, è definito in termini di cambiamenti nei comportamenti e nelle prestazioni dei partecipanti derivanti dall'applicazione sul lavoro degli apprendimenti sviluppati (Kirkpatrick, 1960a). Infine il quarto livello, denominato *results*, mira a determinare se in seguito al percorso vi sono stati miglioramenti negli indicatori di performance a livello organizzativo, ad esempio incrementi nella quantità e qualità dei prodotti/servizi offerti (Kirkpatrick, 1960b).

Sebbene nel frattempo abbia ripreso e riproposto il *framework* all'interno di altri contributi (ad es. Kirkpatrick 1967; 1977), è solo nel 1994 – a oltre trent'anni di distanza dalla sua introduzione – che l'autore ne fornisce una presentazione sistematica, completa e integrata mediante la prima edizione del volume *Evaluating training programs. The four levels*, cui faranno seguito una seconda edizione nel 1998 e una terza nel 2006. Quest'ultima edizione, scritta insieme al figlio James come i due volumi *Transferring learning to behavior* e *Implementing the four levels* pubblicati rispettivamente nel 2005 e nel 2007, presenta il medesimo impianto delle precedenti articolato in due sezioni: a una prima parte incentrata sulla descrizione dei livelli e sulla proposta di relative linee guida nonché di suggerimenti ed esempi, segue una seconda parte contenente alcuni studi di caso che illustrano come i livelli sono stati implementati nell'ambito di differenti programmi e organizzazioni.

Considerando nel complesso i contributi di Kirkpatrick presi in esame, non solo quelli fin qui citati ma anche altri inerenti il modello dei quattro livelli, è possibile cogliere e sottolineare due aspetti che ne costituiscono caratteristiche generali: *a)* il taglio eminentemente divulgativo e pratico degli scritti; *b)* la “costanza” con cui l'autore ha proposto il modello nel corso del tempo.

Per quanto riguarda il primo – riferibile alla presentazione lineare e concisa, al linguaggio accessibile e all'approccio pragmatico che caratterizzano le opere di Kirkpatrick – si tratta di un aspetto non di poco conto, e ciò per almeno due ordini di motivi. Da un lato esso esprime e sottende l'intento primario che ha guidato l'autore nello sviluppo e nella proposta del suo *framework*, nonché il tipo di *audience* privilegiato: come egli stesso esplicitamente dichiara (ad es. Kirkpatrick, 1994; 1996a; 1996b), i suoi scritti si

rivolgono principalmente ai *practitioners* – ovvero ai professionisti della formazione direttamente implicati nella valutazione dei programmi – con l’obiettivo di aiutarli ad affrontare tale compito complesso ma imprescindibile. Dall’altro lato tale aspetto ha avuto un peso rilevante nel favorire la popolarità del modello e dunque, in senso più ampio, nel promuovere la diffusione di pratiche valutative nel mondo della formazione continua. Come affermano vari autori (ad es. Alliger & Janak, 1989; Holton, 1996; Alliger et al., 1997; Tamkin, Yarnall, & Kerrin, 2002; Bates, 2004; Reio, Rocco, Smith, & Chang, 2017), il “potere” del modello e un fattore chiave alla base del suo successo risiedono nella sua semplicità e sobrietà, nella sua capacità di ridurre la complessità della valutazione scomponendola in alcuni concetti e passaggi logici facilmente comprensibili, nell’essere riuscito a intercettare e soddisfare il bisogno di un approccio sistematico ma al contempo chiaro e accessibile.

In relazione al secondo aspetto, è lo stesso Kirkpatrick a confermare, nel 2005, un dato che si ricava dall’analisi in prospettiva diacronica degli scritti presi in esame, prodotti dall’autore nell’arco di circa cinquant’anni: «Mio figlio Jim mi ha chiesto “Quanto è cambiato il tuo modello da quando è stato introdotto nel 1959?”. Gli ho risposto che il modello rimane essenzialmente lo stesso. I concetti, i principi e le tecniche sono applicabili oggi come lo erano quando il modello è stato introdotto per la prima volta» (Kirkpatrick & Kirkpatrick, 2005, p. 4). Anche in un precedente articolo dal titolo *Great ideas revisited. Revisiting Kirkpatrick’s four-level model* (Kirkpatrick, 1996a), che sembrerebbe alludere a un tentativo di revisione/aggiornamento del modello, in realtà si afferma che, nonostante siano passati quasi quarant’anni dalla sua introduzione, «il contenuto è rimasto sostanzialmente lo stesso. Ho apportato alcune modifiche alle linee guida per ciascuno dei quattro livelli, oltre a fornire esempi sempre più numerosi e diversificati nei miei volumi. Ma i quattro livelli sono rimasti i medesimi» (*ibidem*, pp. 54-55). Ciò emerge ad esempio ponendo a confronto il modo in cui vengono definiti i livelli e le relative *guidelines* nella prima serie di articoli degli anni 1959-1960 e, a distanza di quasi cinquant’anni, nella più recente edizione del volume *Evaluating training programs* (Kirkpatrick & Kirkpatrick, 2006): sebbene vi siano alcuni cambiamenti e integrazioni – come l’aggiunta degli atteggiamenti accanto a conoscenze e abilità nell’ambito degli *outcomes* relativi al livello 2, che compare già in un articolo degli anni Settanta (Kirkpatrick, 1977) – l’impianto del modello e le indicazioni generali fornite dall’autore appaiono sostanzialmente invariati nel corso del tempo. Sembra inoltre possibile affermare che neppure il figlio James, pur avendo recentemente proposto in collaborazione con la moglie una versione aggiornata del modello denominata *New World Kirkpatrick Model* (Kirkpatrick & Kayser Kirkpatrick, 2016), sia stato guidato dall’intenzione di rielaborare a fondo la prospettiva valutativa offerta dal padre: le modifiche introdotte nella definizione dei livelli per ampliarne e articolarne il significato, infatti, si inquadrano nel solco dei principi e delle finalità posti alla base del modello originario, di cui si intende anzi mostrare la perdurante rilevanza nel contesto attuale.

Molti sono, in effetti, i “punti forti” del modello di Kirkpatrick, i contributi che ha apportato all’ambito della valutazione della formazione continua. Tra i principali, oltre a quelli già emersi, si possono evidenziare i seguenti (ad es. Alliger & Janak, 1989;

Newstrom, 1995; Holton, 1996; Bates, 2004): ha contribuito a focalizzare la pratica della valutazione su una pluralità di dimensioni ed esiti, ponendo l'esigenza di misure diversificate; operando una chiara distinzione tra *learning* e *behavior*, ha sottolineato l'importanza del processo di *transfer* dell'apprendimento per una reale efficacia della formazione; ha enfatizzato la necessità di pensare e valutare la formazione anche in termini di *business results*, ovvero di ricadute per le aziende; ha avuto una funzione euristica e alimentato lo sviluppo di altri modelli di valutazione come quelli proposti da Hamblin (1974), Kaufman & Keller (1994) e Phillips (1996), che Tamkin et al. (2002) definiscono *descendant models* derivanti dal "modello-padre" di Kirkpatrick.

Tuttavia, pur nel quadro di tale insieme di aspetti positivi che indubbiamente hanno contribuito al suo successo, sembra che la popolarità del modello si sia in un certo senso rivelata un'"arma a doppio taglio", facendogli assumere «un'aura di invulnerabilità» (Newstrom, 1995, p. 317) che ha favorito atteggiamenti di accettazione acritica nonché interpretazioni/applicazioni superficiali e riduttive. Kirkpatrick stesso, del resto, non è mai stato incline a considerare gli "attacchi" sferrati al suo modello: benché non siano mancate rilevanti critiche e obiezioni, l'autore è infatti rimasto fedele alla sua proposta originaria e sembra non aver accolto tali rilievi come opportunità per rivedere e sviluppare il suo approccio. Newstrom (1995) amplia ulteriormente il discorso sostenendo che, oltre a non affrontare le critiche avanzate da altri studiosi, egli nemmeno effettua un'analisi auto-critica dei limiti del suo modello, né esamina il modo in cui può essere comparato o integrato con altri modelli, né infine propone una valutazione critica dei casi di studio presentati nei suoi volumi.

Nel paragrafo che segue, in base a una ricognizione della letteratura sul tema, tenteremo dunque di delineare alcuni rilevanti nodi del dibattito inerente le problematiche connesse al modello di Kirkpatrick e al suo impiego, di cui è importante tener conto non solo per cogliere alcune debolezze dell'approccio in esame da un punto di vista teorico-concettuale, ma anche perché dalla consapevolezza di tali aree di criticità può derivare un uso più valido ed efficace del modello nella pratica.

### Criticità e paradossi di un modello a "legami deboli"

#### *L'uso del modello di Kirkpatrick: un difficile equilibrio tra complessità e fattibilità della valutazione*

Il modello di Kirkpatrick è realmente così conosciuto e utilizzato? Tale interrogativo può sembrare paradossale a fronte del fatto che, come evidenziano pressoché tutti gli autori che l'hanno preso in esame anche con l'intento di sottolinearne limiti e problemi, esso rappresenta ormai da vari decenni il più noto e popolare modello di valutazione della formazione continua.

Ma in che misura tale notorietà si è tradotta in un'effettiva applicazione del modello? Quanto il modo in cui il modello viene utilizzato per valutare i programmi formativi in ambito aziendale rispecchia principi e linee guida definiti dal suo autore? I risultati di alcune ricerche (ad es. Catalanello & Kirkpatrick, 1968; Alliger & Janak, 1989; Plant & Ryan, 1994; Pulichino, 2007; Chang, 2010; Kennedy et al., 2013; Reio et al., 2017) inducono

a ritenere che, dietro al frequente richiamo al modello e alla sua “apparente” diffusione, si cela il più delle volte, nella pratica, un uso riduttivo e improprio. Emergono in particolare due tendenze relative alle prevalenti modalità d’uso del *framework*: da un lato, il focus delle attività valutative si incentra sul primo livello, tutt’al più anche sul secondo, mentre appaiono solo sporadici i tentativi di spingersi fino al terzo e al quarto (3); dall’altro lato, in relazione ai livelli considerati, le prassi valutative attuate sono spesso «superficiali e soggettive» (Catalanello & Kirkpatrick, 1968, p. 9).

Sono dunque proprio i livelli cui Kirkpatrick attribuisce più importanza, ovvero il terzo e il quarto, a delinarsi come gli ambiti maggiormente trascurati e tale riscontro, a fronte di un modello che tra i suoi pregi ha il merito di aver sottolineato l’importanza di valutare la formazione in termini di ricadute organizzative, costituisce un primo elemento di criticità che induce a riflettere sulle ragioni sottostanti: perché, nella maggior parte dei casi, non si va oltre i primi due livelli? Che ostacoli si pongono rispetto alla valutazione degli altri?

Come emerge ad esempio dall’inchiesta di Kennedy et al. (2013), il problema non sembra essere la mancanza di interesse per i livelli “più alti” del modello, anzi, gli autori riscontrano una relazione inversa tra la rilevanza/ utilità attribuita dai *training professionals* ai vari livelli e la frequenza con cui viene valutato ciascuno di essi: per quanto riguarda in particolare il terzo e il quarto, rispettivamente il 98% e l’88% dei rispondenti ritiene importante o molto importante la valutazione di tali livelli, ma solo il 26% e il 13% afferma che essa viene effettivamente condotta nelle rispettive organizzazioni. Emerge insomma un evidente *gap* tra percezioni e prassi che può avere rilevanti effetti sull’efficacia stessa della formazione in quanto – come rivelano i risultati di una ricerca che ha esaminato la relazione tra le modalità di valutazione attuate e l’impatto organizzativo della formazione (Saks & Burke, 2012) – la rilevazione dei cambiamenti nella performance dei lavoratori e dell’azienda mediante attività valutative di *follow-up* appare positivamente associata al *transfer on-the-job*, favorendo la responsabilizzazione degli uni e dell’altra rispetto all’effettiva applicazione nel contesto lavorativo degli apprendimenti sviluppati.

Tornando dunque agli interrogativi sopra formulati, Kennedy et al. (2013) hanno individuato tre fattori chiave, tra loro interrelati, alla base delle difficoltà dei *practitioners* nell’implementazione dei livelli superiori del modello di Kirkpatrick, che appaiono in linea con le problematiche riscontrate da altre ricerche (ad es. Strunk, 1999; Pulichino, 2007; Chang, 2010; Reio et al., 2017). Nel complesso, essi sembrano riconducibili al problema generale della fattibilità della valutazione, a sua volta connesso alla crescente complessità della stessa nel passaggio da un livello all’altro. Il primo e principale ostacolo riguarda i vincoli – in termini di tempi e di risorse umane, materiali e finanziarie – con cui occorre fare i conti nel momento in cui si intende definire/implementare un disegno valutativo; il secondo fattore si riferisce alla mancanza di un supporto adeguato per il processo di valutazione da parte della *leadership* organizzativa, non solo in termini di risorse ma anche di possibilità di accesso ai dati per i valutatori e di “incentivi” per i dipendenti; il terzo fattore è infine costituito dall’insufficiente *expertise* di chi progetta e conduce la valutazione, che solleva l’esigenza di una formazione specifica finalizzata allo sviluppo di conoscenze e competenze valutative.

Degli aspetti critici delineati, come delle conseguenze che ne derivano in termini di uso del modello nella pratica, appare ben consapevole lo stesso Kirkpatrick. Egli sottolinea più volte il problema centrale dei vincoli di fattibilità della valutazione e ciò si riflette nell'“elasticità” con cui sembra intendere le linee guida procedurali proposte per i livelli superiori del modello, aggiungendo ad esempio l'espressione *if practical* (se possibile/fattibile) a fianco di alcune di esse. Inoltre, ai tre fattori sopra considerati, l'autore aggiunge un ulteriore aspetto affermando che «molto tempo e molte energie e risorse vengono dedicati ai livelli 1 e 2 perché questi sono i livelli su cui i professionisti della formazione hanno più controllo» (Kirkpatrick & Kirkpatrick, 2006, p. 83). Come afferma Praslova (2010, p. 220), mentre i primi due livelli del modello sono considerati *interni* in quanto focalizzati su ciò che accade nell'ambito dei percorsi formativi, gli altri si incentrano su cambiamenti che avvengono successivamente e al di fuori dei percorsi, dunque sono visti come criteri *esterni* su cui è probabile influiscano altri fattori di natura contestuale. Ciò rende la valutazione di tali livelli una sfida e spiega inoltre perché i professionisti della formazione credono spesso che «il loro lavoro sia concluso una volta che i programmi formativi finiscono» (Kirkpatrick & Kirkpatrick, 2007, p. 81).

Le problematiche delineate costituiscono, agli occhi dei responsabili della formazione e della sua valutazione, rilevanti ostacoli a una più valida applicazione del modello. Tuttavia, come sostiene Pulichino (2007), il problema del declino dell'uso nel passaggio ai livelli superiori potrebbe avere radici più profonde nella natura e nella struttura stesse dei quattro livelli, rimandando a questioni critiche connesse ai presupposti più o meno esplicitamente associati al modello di Kirkpatrick.

### *Assunti espliciti e impliciti relativi al modello: i legami tra i livelli*

A partire dall'articolo del 1978 di Newstrom, che può essere considerato il primo critico del modello di Kirkpatrick e colui che ha gettato le fondamenta del successivo dibattito, la ricognizione delle critiche avanzate da vari studiosi mostra che il bersaglio principale è rappresentato da alcuni assunti generali relativi al modello – spesso impliciti ma ampiamente diffusi tra i professionisti della formazione e a volte anche tra i ricercatori – considerati potenzialmente dannosi in quanto in grado di indurre chi valuta ad adottare un approccio semplicistico e a trarre conclusioni fuorvianti circa l'efficacia della formazione.

Un primo assunto problematico, che costituisce un punto caldo e controverso del dibattito esaminato (ad es. Newstrom, 1978; Alliger & Janak, 1989; Tamkin et al., 2002; Bates, 2004), concerne l'esistenza di una relazione di causalità lineare/sequenziale o ascendente tra i livelli: si assume cioè che i livelli dipendano l'uno dall'altro, che ciascun livello sia legato al precedente e/o al successivo in una sorta di «catena causale» (Bates, 2004, p. 342). Da questo primo presupposto ne discende un altro (ad es. Newstrom, 1978; Alliger & Janak, 1989) in base al quale ciascun livello sarebbe correlato al precedente e, più in generale, tutte le correlazioni tra i livelli sarebbero positive. Entrambi gli assunti sono ritenuti rischiosi e “ingannevoli” per le loro possibili ripercussioni sulle pratiche valutative: coloro che utilizzano il modello, infatti, possono far propria la convinzione sia che esiti positivi nei livelli più bassi siano sufficienti per desumere esiti altrettanto positivi

in quelli successivi, limitandosi, al limite, a rilevare unicamente le reazioni dei partecipanti; sia che esiti positivi nei livelli più alti implicino risultati altrettanto positivi in quelli precedenti, che possono dunque essere tralasciati per puntare direttamente alla valutazione dell'impatto, dei cambiamenti nelle prestazioni lavorative e nei risultati organizzativi. In entrambi i casi ha luogo una valutazione incompleta, che può portare a inferenze valutative scorrette in merito al valore e all'efficacia complessiva della formazione.

In base a tali premesse, alcuni studi hanno mirato a testare empiricamente le relazioni tra i livelli e la predittività degli esiti relativi ai livelli "inferiori" rispetto agli esiti nei livelli successivi. Ad Alliger e Janak (1989) si deve la prima meta-analisi delle correlazioni tra i livelli di Kirkpatrick, che ha messo in luce un quadro complessivo di legami deboli o del tutto assenti in particolare fra i primi tre livelli: per quanto riguarda il quarto, infatti, gli autori affermano che le associazioni osservate vanno interpretate con estrema cautela dato l'esiguo numero di ricerche che hanno preso in esame tale livello di valutazione. I risultati riportati, che hanno trovato conferma nella successiva e più ampia meta-analisi di Alliger et al. (1997) e in specifiche ricerche empiriche (ad es. Bledsoe, 1999; Fung & Songan, 2016), mostrano innanzitutto che le reazioni, ovvero il criterio di valutazione più frequentemente considerato perché più facilmente rilevabile, sembrano difficilmente utilizzabili «come misure sostitutive dell'apprendimento e del transfer» (Alliger et al., 1997, p. 352). Appare inoltre problematica la relazione tra il secondo e il terzo livello, che rimanda alla complessa questione del *transfer* dell'apprendimento su cui torneremo nel prosieguo del contributo.

Quanto Kirkpatrick stesso può essere ritenuto "responsabile" di aver alimentato gli assunti analizzati, che alla luce delle evidenze empiriche considerate appaiono sostanzialmente infondati? Quanto, nei suoi scritti, implicitamente o esplicitamente assume che i livelli siano tra loro correlati e concatenati mediante legami di natura causale? Non è facile dare una risposta chiara e univoca a tali interrogativi in quanto, nel complesso, i messaggi dell'autore su tale questione sembrano caratterizzati da un certo grado di ambiguità.

Da un lato non si può negare che Kirkpatrick, fin dal momento in cui ha introdotto il suo *framework*, abbia riconosciuto che ciascun livello rappresenta una condizione *necessaria ma non sufficiente* per il livello successivo. A proposito del legame tra i primi due, egli afferma ad esempio che se reazioni positive sono essenziali affinché i partecipanti traggano il massimo beneficio dalla formazione, esse non garantiscono né che l'apprendimento abbia luogo, né che il loro comportamento cambierà in seguito alla formazione (Kirkpatrick, 1959a). In modo ancor più deciso egli riconosce la problematicità del passaggio dal secondo al terzo livello, sostenendo che se l'apprendimento è necessario perché si verifichi un cambiamento nel comportamento occorrono però anche altre condizioni concomitanti, quali la motivazione del soggetto a migliorare le sue prestazioni, l'opportunità effettiva di applicare quanto appreso, un contesto professionale che offra supporto e incoraggiamento (Kirkpatrick, 1960a). Quando poi si entra nel merito del quarto livello il numero dei fattori implicati aumenta ulteriormente, rendendo complicato, a volte impossibile, misurare i risultati organizzativi direttamente attribuibili a specifici



programmi formativi: si tratta del problema della “separazione delle variabili”, cui vengono ricondotte le principali difficoltà connesse alla valutazione di questo livello (Kirkpatrick, 1960b). A questo proposito, in un articolo dal titolo *Evaluating training programs: evidence vs. proof* (Kirkpatrick, 1977), l’autore afferma chiaramente che procedendo dal primo livello ai successivi diventa sempre più difficile ottenere “prove” dell’efficacia della formazione e spesso è necessario accontentarsi di raccogliere “evidenze” dei cambiamenti auspicati. Per produrre “prove”, infatti, occorrono procedure complesse e dispendiose in termini di tempo e di risorse, che richiedono l’adozione di un «un approccio di ricerca» (*ibidem*, p. 11) basato su un disegno sperimentale.

D’altra parte, tuttavia, nei contributi di Kirkpatrick é possibile rintracciare varie affermazioni che – in maniera più implicita nei primi scritti e poi progressivamente più esplicita – sembrano dare adito agli assunti fatti propri dai *practitioners* e identificati dai critici, implicando l’esistenza di legami di causalità lineare tra i livelli (ad es. Newstrom, 1978, 1995; Holton, 1996; Pulichino, 2007). Alcuni passaggi indicativi in tal senso ci sembrano ad esempio individuabili nel volume del 2007 *Implementing the four levels*, in cui compare anche il concetto di “catena di evidenze” – poi ripreso dal figlio James in altri contributi successivi (ad es. Kirkpatrick & Kayser Kirkpatrick, 2009) – che in questa prospettiva di analisi appare particolarmente rilevante. Il senso attribuito a tale espressione ci sembra riassumibile nella seguente affermazione: «Se hai fatto un buon lavoro nell’assicurare un ambiente formativo favorevole all’apprendimento (e hai evidenze per dimostrarlo), e se puoi documentare che gli obiettivi di apprendimento sono stati raggiunti, portando poi ai comportamenti lavorativi desiderati, allora i risultati relativi al livello 4 [...] dovrebbero provare questa relazione fantastica e convincente tra formazione e risultati aziendali» (Kirkpatrick & Kirkpatrick, 2007, p. 128). Il termine “catena”, afferma Kirkpatrick, è fondamentale: esso implica non solo che ciascun livello è collegato al successivo, ma anche che il processo di valutazione deve rispettare la sequenza dei livelli, in un continuum di passi da compiere in successione e senza saltare nessuna tappa. Percorrendo tale sequenza, vale a dire raccogliendo un opportuno *mix* di dati per ciascun livello a cominciare dal primo, è possibile ricostruire la catena degli eventi e raccontare la «storia del valore della formazione» (*ibidem*, p. 83), dimostrando il suo contributo in relazione alle attese degli *stakeholders*: il concetto di *chain of evidence* può infatti essere definito come insieme di dati, informazioni e testimonianze relativi a ciascuno dei quattro livelli che, quando presentati in sequenza, dimostrano il valore della formazione per l’organizzazione (Kirkpatrick & Kayser Kirkpatrick, 2009). In modo simile a ciò che fa un buon avvocato per preparare la sua arringa finale, se il valutatore costruisce tale catena raccogliendo indizi sufficienti a determinare una convincente successione di legami di «causa-effetto» (*ibidem*, p. 187), porterà la giuria a esprimere «un verdetto positivo: un verdetto in cui gli *stakeholders* affermano che, oltre ogni ragionevole dubbio, l’intervento formativo ha contribuito con successo ai risultati attesi» (*ibidem*, p. 173).

Le perplessità relative alla prospettiva delineata sono riconducibili a diversi aspetti. Da un lato, considerando anche le deboli correlazioni riscontrate tra i livelli, rimane aperta una serie di questioni relative alla possibilità di ottenere “dati misti” ovvero dati che rivelano esiti contrastanti tra i diversi livelli di valutazione o all’interno dei singoli livelli

(Newstrom, 1978; 1995). Dall'altro lato, anche nell'eventualità di ottenere esiti complessivamente positivi a tutti i livelli, ciò risulta sufficiente per dimostrare "oltre ogni ragionevole dubbio" il valore della formazione? In altre parole, se si osservano esiti positivi a un dato livello, è sufficiente rilevare esiti positivi anche ai livelli precedenti per affermare che i primi *dipendono* dai secondi? In base al concetto di "catena di evidenze", la risposta di Kirkpatrick e del figlio parrebbe positiva. Ciò appare, tuttavia, sostanzialmente in contraddizione con quanto affermato dall'autore circa la differenza tra "prova" ed "evidenza" nell'articolo del 1977 precedentemente citato, sollevando inoltre questioni di ordine metodologico relative alla qualità delle procedure e degli strumenti di valutazione. Come abbiamo visto, infatti, Kirkpatrick associa la possibilità di produrre vere e proprie "prove" dell'efficacia della formazione all'uso di rigorosi disegni sperimentali, mentre in relazione alla costruzione di "evidenze" non sembrano esservi specifici standard metodologici a garanzia della validità e attendibilità delle informazioni raccolte. Si afferma anzi che, in generale, il livello di complessità e formalità della catena di evidenze dipenderà dalle dimensioni e dall'importanza dell'iniziativa di formazione valutata, suggerendo un approccio in cui il valutatore è chiamato a fare il meglio che può a seconda delle situazioni e dei vincoli che di volta in volta si presentano (Kirkpatrick & Kirkpatrick, 2007).

In sostanza, ci sembra di poter concordare con Holton (1996) quando sostiene che Kirkpatrick non è chiaro sulla questione dei legami causali all'interno del suo modello: da una parte egli riconosce che vari fattori possono influenzare gli *outcomes* della formazione e dunque che le relazioni tra i livelli non sono semplici e lineari; d'altra parte però propone affermazioni e concetti da cui sembra possibile inferire una semplice relazione di causalità sequenziale tra i livelli, mentre per trarre conclusioni di natura causale occorre, afferma Holton, sviluppare modelli di valutazione più completi e complessi.

Prima di esaminare nel successivo paragrafo la proposta dell'autore in tale direzione, vorremmo soffermarci su un ulteriore assunto che a differenza dei precedenti appare codificato nel modello di Kirkpatrick in modo del tutto esplicito e intenzionale. In base a tale assunto, il modello implica una gerarchia di valore tra i vari livelli, per cui procedendo dal primo al quarto «la valutazione diventa più difficile e costosa» ma anche «più importante e significativa» (Kirkpatrick 1996a, p. 56). L'autore è chiaro su questo punto: le informazioni più rilevanti per valutare l'efficacia di un programma formativo sono quelle relative al livello 4, che incarna i fini ultimi della formazione. Egli afferma ad esempio (Kirkpatrick & Kirkpatrick, 2006, p. 59): «È importante capire che il cambiamento nel comportamento non rappresenta un fine in sé. Piuttosto, è un mezzo per un fine: i risultati finali che possono essere raggiunti se si verifica un cambiamento nel comportamento», sostenendo inoltre che il livello 4 «è il più importante per i nostri *stakeholders* (i nostri partner aziendali). Pertanto, deve essere il più importante per noi (come professionisti della formazione)» (Kirkpatrick & Kirkpatrick, 2007, p. 107).

Il presupposto che le quattro dimensioni proposte costituiscano livelli gerarchici di *outcomes* della formazione, posti in ordine ascendente in base al valore delle informazioni che forniscono, può essere considerato problematico sia perché può indurre a trascurare i primi livelli e a sottovalutarne la rilevanza, sia perché veicola l'idea che ciò che conta sia il

punto di vista dell'azienda (ad es. Newstrom, 1978; Alliger & Janak, 1989; Bates, 2004). Rispetto a tale questione e alle potenziali criticità connesse all'adozione di un modello di valutazione "azienda-centrico", Stokking (1996) suggerisce che sarebbe più opportuno parlare di *criteri* piuttosto che di *livelli* di valutazione, evitando così di utilizzare un termine che evoca un ordine di importanza tra le categorie considerate.

### *Un modello o una tassonomia? Variabili intervenienti e dimensioni processuali*

Il modello di Kirkpatrick è effettivamente un modello? Anche questo interrogativo, come quello preso in esame all'inizio del secondo paragrafo, può apparire paradossale: il *framework* proposto è infatti comunemente noto come "modello dei quattro livelli" e in tal modo viene generalmente definito anche nella letteratura scientifica inerente la valutazione della formazione continua. Alcuni autori hanno tuttavia sottolineato che esso poggia su deboli e insicure fondamenta teoriche, che non è stato sufficientemente validato dalla ricerca, fino a mettere in discussione il fatto stesso che si tratti realmente di un modello.

Nella sua recensione alla prima edizione del volume *Evaluating training programs*, Newstrom sostiene che il lavoro di Kirkpatrick è «totally atheoretical» (Newstrom, 1995, p. 319), basato su presupposti per lo più impliciti, poco specificati e riduttivi, privo di riferimenti al ruolo della ricerca nel produrre conoscenza sulla valutazione. Il suo volume non fornisce dunque, ad avviso dell'autore, «una solida spiegazione basata sulla teoria delle multiple dimensioni e delle complesse questioni relative alla valutazione di programmi e attività di sviluppo professionale» (*ibidem*, p. 320). McLean e Moss (2003) constatano che, mentre nella letteratura sulla formazione continua si fa sistematicamente riferimento al *framework* di Kirkpatrick inteso come modello, ciò accade meno frequentemente nella letteratura più generale sulla valutazione dei programmi: Stufflebeam (2001), ad esempio, non fa riferimento a Kirkpatrick nella sua descrizione e valutazione di 22 approcci o modelli di valutazione dei programmi. Gli autori affermano inoltre che, «per poter essere incluso in un elenco di modelli generali di valutazione dei programmi, il framework di Kirkpatrick deve essere ampliato» (McLean e Moss, 2003, p. 20). Anche altri autori (ad es. Warr & Bunce, 1995; Warr, Allan, & Birdi, 1999; Bates, 2004) hanno individuato un punto debole del modello nel suo essere incompleto e riduttivo, nel suo proporre un approccio semplicistico in cui l'analisi del sistema di influenze che agiscono sull'efficacia della formazione sembra non costituire un elemento essenziale del processo di valutazione. Il modello viene infatti generalmente ricondotto a un approccio valutativo di stampo positivista-sperimentale, centrato sulla misurazione degli *outcomes* della formazione e sul confronto obiettivi-risultati (ad es. Eseryel, 2002; Bartezzaghi, Guerci, & Vinante, 2010; Coldwell & Simkins, 2011), in cui la valutazione ha luogo quando il programma è concluso per determinarne il merito e il valore assumendo una funzione prevalentemente sommativa (ad es. Newstrom, 1995; Reio et al., 2017). Questa centratura sui prodotti, sulla quantificazione degli esiti della formazione – cui si accompagna un'enfasi sull'*accountability* come finalità principale, anche se non esclusiva (4), della valutazione – porta a trascurare la dimensione dei processi sottostanti, dunque a inibire la

comprensione delle dinamiche che contribuiscono a determinare i risultati che si osservano.

Holton, che può forse essere considerato come “il più critico tra i critici” del modello di Kirkpatrick, elenca le sei componenti che una teoria o un modello dovrebbero possedere in base ai criteri definiti da Klimoski (1991), poi afferma che esso non ne soddisfa nessuno. Già considerando la prima delle componenti elencate, ovvero un insieme completo dei costrutti essenziali sul fenomeno in esame, appare evidente l’assenza nel modello di Kirkpatrick delle principali variabili intervenienti – riconducibili a fattori individuali e di contesto – in grado di influenzare gli esiti ai vari livelli: nessun modello di valutazione, sostiene l’autore, può essere validato senza misurare e controllare gli effetti di queste variabili. Da qui l’idea che il *framework* proposto da Kirkpatrick non costituisca un modello ma semplicemente una tassonomia, uno schema di descrizione e classificazione di *outcomes* della formazione continua; non fornisca un modello di valutazione sistemico, ma solo quattro distinti e indipendenti criteri di valutazione. Un limite delle tassonomie, afferma l’autore, è che non identificano il complesso insieme di costrutti sottostanti al fenomeno di interesse, dunque non sorprende che le presunte relazioni di causalità lineare tra i livelli non siano state dimostrate dalla ricerca: infatti, «per definizione, le tassonomie classificano piuttosto che definire costrutti causali» (Holton, 1996, p. 6).

Sottolineando la necessità di irrobustire la ricerca nel settore per andare oltre un approccio tassonomico, orientandosi verso un modello più completo e articolato di valutazione della formazione continua, Holton (1996, 2005) propone un passo in tale direzione presentando un *framework* concettuale sviluppato in base all’analisi e integrazione delle evidenze emerse dalla letteratura sull’argomento, in particolare dagli studi sul processo di *transfer* dell’apprendimento nel contesto professionale. La questione del *learning transfer* o *training transfer* è stata infatti ampiamente esplorata soprattutto dagli anni Novanta del secolo scorso, divenendo oggetto di numerose ricerche volte a indagare il ruolo dei fattori implicati in tale fenomeno multidimensionale e complesso. L’interesse per il tema è testimoniato anche dalla disponibilità di diverse rassegne della letteratura (per citare solo alcune tra le più recenti: Merriam & Leahy, 2005; Devos & Dumay, 2006; Burke & Hutchins, 2007; Cheng & Hampson, 2008; Blume, Ford, Baldwin, & Huang, 2010; Khan, Mufti, & Nazir, 2015), dalle quali emerge che il *transfer* è influenzato da un insieme di aspetti complessivamente riconducibili a tre macro-dimensioni: la dimensione “individuo”, relativa a caratteristiche dei partecipanti come la motivazione e l’auto-efficacia; la dimensione “formazione”, relativa a caratteristiche dei percorsi quali l’attenzione posta sull’aspetto applicativo dell’apprendimento e il riferimento ad attività/ situazioni simili a quelle che i partecipanti incontrano o possono incontrare nel proprio lavoro; la dimensione “contesto professionale”, che include ad esempio l’effettiva opportunità di utilizzare sul lavoro quanto appreso ma anche costrutti complessi come quelli di *transfer climate* e *learning culture*, sostanzialmente riferibili al tipo di supporti e incentivi che i partecipanti ricevono – da superiori e colleghi – nel contesto lavorativo.

Il modello proposto da Holton (cfr. Figura 1) si differenzia da quello di Kirkpatrick per due aspetti principali. In primo luogo, i tipi di *outcomes* considerati sono tre e non quattro: cambia infatti il ruolo delle cosiddette *reactions*, non più viste come un *outcome* della

formazione legato all'apprendimento da una relazione lineare ma piuttosto come una variabile moderatrice che – in interazione con altre variabili – può influenzare l'apprendimento e venirne a sua volta influenzata. In secondo luogo, tenendo conto di una molteplicità di fattori in grado di incidere sull'efficacia della formazione, il modello include e specifica le relazioni tra diverse variabili intervenienti distinte in primarie – influenzano direttamente gli *outcomes* e sono ricondotte a tre principali ambiti che rispecchiano le tre macro-dimensioni sopra citate – e secondarie – influenzano le variabili primarie, dunque indirettamente gli *outcomes*.

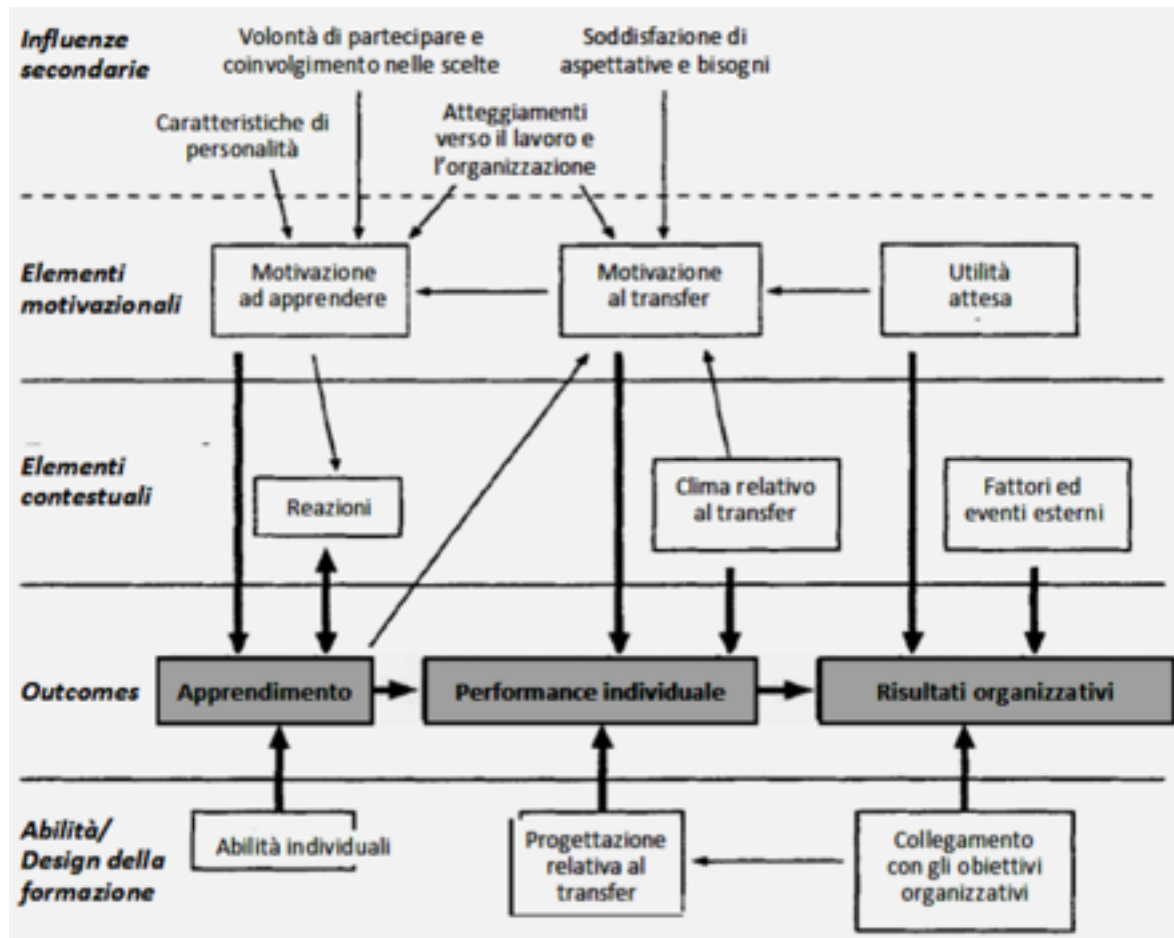


FIG. 1 - HRD EVALUATION RESEARCH AND MEASUREMENT MODEL (ADATTAMENTO DA HOLTON, 1996)

In relazione alla questione affrontata, ovvero se il *framework* di Kirkpatrick costituisca o meno un modello, è necessario puntualizzare che, almeno inizialmente, l'autore non l'ha presentato come tale. Nelle sue prime pubblicazioni Kirkpatrick non usa la parola "modello" e neppure parla di "livelli" di valutazione – utilizzando piuttosto termini quali "step", "stadi", "segmenti" o "criteri" (Kirkpatrick, 1959a; 1967; 1977) – e quando inizia a utilizzare tali espressioni lo fa sostanzialmente accogliendo le definizioni proposte da altri e affermatesi nell'uso comune (ad es. Kirkpatrick, 1996a). Come precedentemente sottolineato, egli offre una prospettiva pratica piuttosto che teorica e, in quest'ottica, la

dura critica di Holton è ritenuta marginale rispetto al suo intento e valore di fondo: «Personalmente, non mi interessa se il mio lavoro viene chiamato modello o tassonomia, purché contribuisca a chiarire in termini semplici il significato della valutazione e offra linee guida e suggerimenti su come condurre una valutazione» (Kirkpatrick, 1996b, pp. 23-24) e «purché i professionisti della formazione lo trovino utile per valutare i programmi formativi» (Kirkpatrick, 1996a, p. 55). Ciò che conta, insomma, è che il modello sia di aiuto ai *practitioners*, ai quali «non interessa un approccio scientifico e complesso. Vogliono qualcosa che possono capire e utilizzare» (*ibidem*). Egli scrive inoltre: «Nel mondo reale in cui avviene la valutazione della formazione, la parola “modello” è comunemente usata per descrivere un modo sistematico di fare qualcosa. [...] La parola “modello” sembra comunicare che si tratta di qualcosa che può essere utilizzato come un’utile guida» (Kirkpatrick, 1996b, pp. 23-24).

Sembra dunque che Kirkpatrick, nel fare proprio il termine “modello” comunemente utilizzato per definire il suo *framework* valutativo, faccia riferimento a un’accezione del termine in senso *prescrittivo* più che *descrittivo/esplicativo*: come afferma Christie (2012) riprendendo la distinzione proposta da Alkin e Ellett (1990), mentre i modelli del primo tipo mirano a guidare la pratica piuttosto che a spiegare i fenomeni, offrendo principi logici e organizzativi per le scelte procedurali fatte dai valutatori e specificando come vada condotta una valutazione buona o corretta, i modelli del secondo tipo sono teorie della valutazione fondate su basi empiriche, più vicini alla tradizionale nozione di teoria nelle scienze sociali e interessati alla valutazione come disciplina.

## Conclusioni

Il presente contributo non ha inteso mettere in dubbio il valore e l’utilità del modello proposto da Donald Kirkpatrick per la valutazione della formazione continua. Piuttosto, analizzando il modello in una prospettiva storica e critica, si è cercato di ricostruire alcune rilevanti aree di complessità e criticità di cui è importante avere consapevolezza per non farsi attrarre e “ingannare” dal richiamo della sua apparente semplicità, che indubbiamente ha contribuito alla sua grande popolarità. Sulla base della disamina effettuata ripercorrendo gli scritti dell’autore in connessione agli sviluppi del dibattito relativo al modello, è possibile affermare che alcune rilevanti problematiche emergono sia relativamente al modo in cui il *framework* dei quattro livelli viene utilizzato nella pratica, sia in riferimento agli assunti sottostanti. Dietro ad applicazioni spesso improprie e riduttive, infatti, sono stati individuati problemi legati non solo alla mancanza di adeguate risorse e competenze per condurre la valutazione, ma anche a modalità generali, diffuse e spesso implicite di interpretare il modello – che Kirkpatrick stesso sembra aver contribuito ad alimentare – messe in discussione da vari studiosi e dagli esiti di specifiche ricerche empiriche.

In conclusione si propongono dunque alcune riflessioni in merito a possibili implicazioni derivanti dall’analisi proposta sul piano della ricerca e della pratica valutativa: gli aspetti richiamati, infatti, costituiscono possibili direzioni di azione per un uso del modello più consapevole e critico e, al contempo, possibili prospettive di sviluppo

e di approfondimento per la ricerca nel campo della valutazione della formazione continua.

Un primo punto di attenzione è relativo alla necessità di considerare e comprendere la complessa rete di fattori che entrano in gioco nei processi formativi e contribuiscono a determinarne l'efficacia e l'impatto: se le categorie di *outcomes* proposte da Kirkpatrick rimangono un utile riferimento per guidare la valutazione dei programmi, esse vanno tuttavia inquadrare in una più ampia analisi delle dimensioni e dinamiche implicate al fine di trarre conclusioni più valide sugli effetti riconducibili agli interventi formativi. Il modello proposto da Holton preso in esame nel secondo paragrafo è stato elaborato per rispondere a tale esigenza, inserendo la misurazione degli *outcomes* all'interno di un approccio valutativo sistemico e multidimensionale in cui si considerano le relazioni tra tali esiti e molteplici fattori di natura individuale e contestuale. Ciò induce inoltre a sottolineare il ruolo cruciale della ricerca sia nel contribuire alla definizione/operazionalizzazione dei costrutti in esame, sia nel mettere a punto metodi e strumenti per rilevare e valutare tali costrutti, con particolare riferimento agli ultimi due livelli del modello di Kirkpatrick su cui, come abbiamo visto, le conoscenze e competenze dei *practitioners* appaiono più deboli. Il *Learning Transfer System Inventory* costruito e validato da Holton e collaboratori (Holton, Bates, & Ruona, 2000) costituisce un esempio in tal senso.

Strettamente associato al primo è il secondo aspetto su cui ci soffermiamo, anch'esso riferibile alla necessità di promuovere una prospettiva valutativa di natura sistemica. La comprensione dell'intricata rete di influenze che agiscono sugli *outcomes* della formazione continua, infatti, è importante non solo per valutare in modo più rigoroso e attendibile gli esiti finali, ma anche per sapere su quali leve si può agire per far sì che tali esiti siano positivi e soddisfacenti. In altre parole, tali influenze non costituiscono solo variabili di "disturbo" da controllare per formulare valide inferenze sui risultati della formazione, ma anche elementi essenziali da considerare quali fattori di *input* e di processo nelle fasi di progettazione e implementazione dei programmi. Ciò richiede una visione più ampia delle funzioni e del ruolo della valutazione nei processi formativi, capace di superare i limiti dell'approccio di Kirkpatrick riconducibili alla sua centratura sul confronto obiettivi-risultati in un'ottica sommativa e rendicontativa. In tale direzione sembrano andare ad esempio le proposte formulate da alcuni autori come Khalid, Rehman e Ashraf (2012), che hanno elaborato un *framework* per la valutazione della formazione continua nato da un tentativo di integrazione tra il modello dei quattro livelli di Kirkpatrick e il modello CIPP di Stufflebeam, e Pineda (2010), che ha integrato il modello di Kirkpatrick aggiungendovi una dimensione definita *pedagogical coherence/adequacy* dei processi formativi, ritenuta essenziale per incentivare la «*optimising function*» della valutazione. Trascurando una dimensione chiave della valutazione – quella *pedagogica*, che consente di migliorare la qualità dei processi – Kirkpatrick ha infatti dato insufficiente rilievo al suo uso in senso formativo-regolativo, che risulta fondamentale soprattutto per i programmi più complessi e di più lunga durata.

Agli aspetti considerati si connette un'altra questione, accennata nell'ambito del contributo, relativa alla necessità che i professionisti della formazione ridefiniscano il

modo in cui a volte concepiscono il proprio ruolo e la propria responsabilità rispetto agli *outcomes* della formazione e alla loro valutazione. Si tratta di un aspetto che appare particolarmente rilevante in riferimento al complesso e dibattuto problema del *transfer* dell'apprendimento, la cui valutazione suscita a volte ansie, perplessità e resistenze legate anche al fatto che solo in parte ci si sente responsabili degli esiti che si producono al di fuori dell'aula. La molteplicità dei fattori che incidono sul processo dovrebbe invece condurre all'adozione di una "strategia sistematica/sistemica di gestione del *transfer*" (Lim, Yoon, & Park, 2013), che intervenga sulle variabili critiche prima, durante e dopo l'esperienza formativa al fine di promuovere gli esiti finali auspicati. Ciò significa assumersi una responsabilità *diretta* per tali esiti, ponendosi il problema del *transfer* non solo in fase di valutazione *ex post* ma già in fase di definizione degli obiettivi e delle modalità attuative dei percorsi, e *indiretta*, cercando ad esempio di cogliere i potenziali ostacoli al *transfer* nel contesto lavorativo dei partecipanti e di contribuire a prevenirli e/o ad affrontarli.

Un ulteriore punto critico da considerare rispetto all'uso del modello è infine la sua centratura – chiaramente espressa dalla gerarchia di valore posta tra gli esiti relativi ai vari livelli – sul punto di vista dell'impresa/organizzazione intesa quale sistema di interessi unitario. Come affermano Bartezzaghi et al. (2010), nel modello manca una prospettiva multi-attore e si trascura la pluralità degli interessi e delle esigenze potenzialmente conflittuali che i diversi *stakeholders* coinvolti, interni ed esterni all'organizzazione, possono avere in merito alla formazione e alla sua valutazione. Se lo stesso Kirkpatrick sottolinea la necessità di una *partnership* con il *management* aziendale è importante però adottare una prospettiva più ampia, negoziata e partecipata in merito agli *stakeholders* da coinvolgere nella definizione dei percorsi, dei relativi piani di valutazione e anche delle modalità di comunicazione/distribuzione/uso dei dati raccolti, che responsabilizzi tutti gli attori chiave e promuova la loro adesione/collaborazione al processo di valutazione – condizione essenziale per la sua fattibilità – nonché l'uso effettivo delle informazioni che esso è in grado di fornire.

In sintesi, e in conclusione, le prospettive di azione e di approfondimento individuate appaiono in linea con quanto sostengono Coldwell e Simkins (2011) quando, prendendo in esame il modello di Kirkpatrick, affermano che non rende giustizia alla complessità dei processi e degli effetti implicati nella formazione continua, la quale richiede un *range* di approcci valutativi piuttosto che il riferimento a un singolo modello «quantitativo, unitario e strumentale» (*ibidem*, p. 154). La consapevolezza dei limiti e delle criticità del modello diventa dunque condizione essenziale per effettuare in relazione al suo uso scelte teoricamente fondate che, scardinando abiti valutativi consolidati, aprano il varco a nuove possibilità di integrazione con approcci più sistemici, plurali e orientati alla comprensione.

## Note

- (1) Tutte le citazioni dirette riportate in questo contributo, tratte dagli scritti di Kirkpatrick ma anche da articoli e opere di altri studiosi, sono state tradotte dall'autrice.



- (2) In ambito accademico, ad esempio, il modello di Kirkpatrick è stato individuato quale utile riferimento per valutare l'efficacia sia dei programmi e dei corsi di studio (ad es. Praslova, 2010; Aryadoust, 2016), sia dei percorsi di formazione e sviluppo professionale rivolti ai docenti universitari (ad es. Giovannini & Rosa, 2012).
- (3) Dal rapporto del 2007 *State of the Industry Report* dell'*American Society for Training and Development*, ad esempio, emerge che mentre la quasi totalità dei programmi formativi (94%) viene valutata al livello 1, le percentuali si riducono sensibilmente nel passaggio al livello 2 (34%) e ancor più ai livelli 3 (13%) e 4 (3%) (Chang, 2010). Pulichino (2007), nella sua inchiesta su un ampio campione di *training professionals*, rileva percentuali più elevate per i livelli 2 (56%), 3 (20%) e 4 (13%), ma sostanzialmente i suoi dati confermano il medesimo andamento decrescente rispetto all'uso dei quattro livelli di valutazione.
- (4) Nei suoi volumi (ad es. Kirkpatrick & Kirkpatrick, 2006) l'autore assegna alla valutazione anche il compito di fornire informazioni utili a migliorare la qualità dei programmi futuri, affermando inoltre che il modello può essere utilmente impiegato in fase di progettazione e sviluppo dei programmi attraverso un percorso "a ritroso" – dal quarto livello al primo – rispetto a quello da attuare in fase di valutazione finale.

## Bibliografia

- Alkin, M.C., & Ellett, F.S. (1990). Development of evaluation models. In H.J. Walberg & G.D. Haertel (Eds.), *The international encyclopedia of educational evaluation* (pp. 15-21). Oxford, UK: Pergamon Press.
- Alliger, G.M., & Janak, E.A. (1989). Kirkpatrick's levels of training criteria: thirty years later. *Personnel Psychology*, 42(2), 331-342.
- Alliger, G.M., Tannenbaum, S.I., Bennett, W., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology*, 50(2), 341-358.
- Aryadoust, V. (2016). Adapting Levels 1 and 2 of Kirkpatrick's model of training evaluation to examine the effectiveness of a tertiary-level writing course. *Pedagogies: An International Journal*, 12(2), 151-179.
- Bartezzaghi, E., Guerci, M., & Vinante, M. (2010). *La valutazione stakeholder-based della formazione continua. Modelli, processi, strumenti*. Milano: Franco Angeli.
- Bates, R. (2004). A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence. *Evaluation and Program Planning*, 27(3), 341-347.
- Bledsoe, M. (1999). *Correlations in Kirkpatrick's training evaluation model*. ProQuest LLC, UMI Dissertation Publishing, Number: 9937553.
- 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.
- Burke, L.A., & Hutchins, H.M. (2007). Training transfer: an integrative literature review. *Human Resource Development Review*, 6(3), 263-296.
- Catalanello, R.F., & Kirkpatrick, D.L. (1968). Evaluating training programs: the state of the art. *Training and Development Journal*, 22(5), 2-9.
- Chang, Y.H.E. (2010). *An empirical study of Kirkpatrick's evaluation model in the hospitality industry*. ProQuest LLC, UMI Dissertation Publishing, Number: 3447779.
- Cheng, E.W., & Hampson, I. (2008). Transfer of training: a review and new insights. *International Journal of Management Reviews*, 10(4), 327-341.
- Christie, C.A. (2012). Advancing empirical scholarship to further develop evaluation theory and practice. *The Canadian Journal of Program Evaluation*, 26(1), 1-18.
- Coldwell, M., & Simkins, T. (2011). Level models of continuing professional development evaluation: a grounded review and critique. *Professional Development in Education*, 37(1), 143-157.
- Devos, C., & Dumay, X. (2006). Les facteurs qui influencent le transfert: une revue de la littérature. *Savoirs*, 12(3), 11-46.
- Eseryel, D. (2002). Approaches to evaluation of training: theory & practice. *Educational Technology and Society*, 5(2), 93-98.
- Fung, C.S., & Songan, P. (2016). The dynamics of training outcomes in a training experience. *Journal of Cognitive Sciences and Human Development*, 1(2), 34-45.
- Giovannini, M.L., & Rosa, A. (2012). La valutazione di impatto dei progetti di formazione all'insegnamento dei docenti universitari: quali indicazioni dalle rassegne delle ricerche empiriche? *Giornale Italiano della Ricerca Educativa*, 8, 93-104.

- Hamblin, A.C. (1974). *Evaluation and control of training*. New York, NY: McGraw-Hill.
- Holton III, E.F. (1996). The flawed four-level evaluation model. *Human Resource Development Quarterly*, 7(1), 5-21.
- Holton III, E.F., Bates, R.A., & Ruona, W.E.A. (2000). Development of a generalized Learning Transfer System Inventory. *Human Resource Development Quarterly*, 11(4), 333-360.
- Holton III, E.F. (2005). Holton's evaluation model: new evidence and construct elaborations. *Advances in Developing Human Resources*, 7(1), 37-54.
- Kaufman, R., & Keller, J.M. (1994). Levels of evaluation: beyond Kirkpatrick. *Human Resource Development Quarterly*, 5(4), 371-380.
- Kennedy, P.E., Chyung, S.Y., Winiecki, D.J., & Brinkerhoff, R.O. (2013). Training professionals' usage and understanding of Kirkpatrick's Level 3 and Level 4 evaluations. *International Journal of Training and Development*, 18(1), 1-21.
- Khalid, M.M., Rehman, C.A., & Ashraf, M. (2012). Exploring the link between Kirkpatrick (KP) and context, input, process and product (CIPP) training evaluation models. *African Journal of Business Management*, 6(1), 274-279.
- Khan, I., Mufti, S., & Nazir, N.A. (2015). Transfer of training: a reorganized review on work environment and motivation to transfer. *International Journal of Management, Knowledge and Learning*, 4(2), 197-219.
- Kirkpatrick, D.L. (1959a). Techniques for evaluating training programs - Part 1: Reaction. *Journal of the American Society of Training Directors*, 13(11), 3-9.
- Kirkpatrick, D.L. (1959b). Techniques for evaluating training programs - Part 2: Learning. *Journal of the American Society of Training Directors*, 13(12), 21-26.
- Kirkpatrick, D.L. (1960a). Techniques for evaluating training programs - Part 3: Behavior. *Journal of the American Society of Training Directors*, 14(1), 13-18.
- Kirkpatrick, D.L. (1960b). Techniques for evaluating training programs - Part 4: Results. *Journal of the American Society of Training Directors*, 14(2), 28-32.
- Kirkpatrick, D.L. (1967). Evaluation of training. In R.L. Craig & L.R. Bittel (eds.), *Training and Development Handbook* (pp. 87-112). New York, NY: McGraw-Hill.
- Kirkpatrick, D.L. (1977). Evaluating training programs: evidence vs. proof. *Training and Development Journal*, 31(11), 9-12.
- Kirkpatrick, D.L. (1994). *Evaluating training programs. The four levels* (1st ed.). San Francisco, CA: Berrett-Koehler.
- Kirkpatrick, D.L. (1996a). Great ideas revisited. Revisiting Kirkpatrick's four-level model. *Training & Development*, 50(1), 54-59.
- Kirkpatrick, D.L. (1996b). Invited reaction: Reaction to Holton article. *Human Resource Development Quarterly*, 7(1), 23-25.
- Kirkpatrick, D.L. (1998). *Evaluating training programs. The four levels* (2nd ed.). San Francisco, CA: Berrett-Koehler.
- Kirkpatrick, D.L., & Kirkpatrick, J.D. (2005). *Transferring learning to behavior. Using the four levels to improve performance*. San Francisco, CA: Berrett-Koehler.

- Kirkpatrick, D.L., & Kirkpatrick, J.D. (2006). *Evaluating Training Programs. The Four Levels* (3rd ed.). San Francisco, CA: Berrett-Koehler.
- Kirkpatrick, D.L., & Kirkpatrick, J.D. (2007). *Implementing the four levels. A practical guide for effective evaluation of training programs*. San Francisco, CA: Berrett-Koehler.
- Kirkpatrick, J.D., & Kayser Kirkpatrick, W. (2009). *Kirkpatrick then and now. A strong foundation for the future*. Saint Louis, MO: Kirkpatrick Partners, LLC.
- Kirkpatrick, J.D., & Kayser Kirkpatrick, W. (2016). *Kirkpatrick's four levels of training evaluation*. Alexandria, VA: ATD Press.
- Klimoski, R (1991). Theory presentation in human resource management. *Human Resource Management Review*, 1(4), 253-271.
- Lim, D.H., Yoon, S.W., & Park, S. (2013). Integrating learning outcome typologies for HRD: review and current status. *New Horizons in Adult Education and Human Resource Development*, 25(2), 33-48.
- Marcuccio, M. (2012). *Valutare l'impatto della formazione continua nelle imprese. Scelte teoriche e operative*. Parma: Universitas.
- McLean, S., & Moss, G. (2003). They're happy, but did they make a difference? Applying Kirkpatrick's framework to the evaluation of a national leadership program. *Canadian Journal of Program Evaluation*, 18(1), 1-23.
- Merriam, S.B., & Leahy, B. (2005). Learning transfer: a review of the research in adult education and training. *PAACE Journal of Lifelong Learning*, 14(1), 1-24.
- Newstrom, J.W. (1978). Catch-22: The problems of incomplete evaluation of training. *Training and Development Journal*, 32(11), 22-24.
- Newstrom, J.W. (1995). Review of "Evaluation training programs. The four levels". *Human Resource Development Quarterly*, 6(3), 317-320.
- Phillips, J.J. (1996). ROI: The search for best practices. *Training & Development*, 50(2), 42-47.
- Pineda, P. (2010). Evaluation of training in organizations: a proposal for an integrated model. *Journal of European Industrial Training*, 34(7), 673-693.
- Plant, R.A., & Ryan, R.G. (1994). Who is evaluating training? *Journal of European Industrial Training*, 18(5), 27-30.
- Praslova, L. (2010). Adaptation of Kirkpatrick's four level model of training criteria to assessment of learning outcomes and program evaluation in Higher Education. *Educational Assessment, Evaluation and Accountability*, 22(3), 215-225.
- Pulichino, J.P. (2007). *Usage and value of Kirkpatrick's four levels of training evaluation*. ProQuest LLC, UMI Dissertation Publishing, Number: 3264775.
- Reio, T.G., Rocco, T.S., Smith, D.H., & Chang, E. (2017). A critique of Kirkpatrick's evaluation model. *New Horizons in Adult Education and Human Resource Development*, 29(2), 35-53.
- Saks, A.M., & Burke, L.A. (2012). An investigation into the relationship between training evaluation and the transfer of training. *International Journal of Training and Development*, 16(2), 118-127.
- Stokking, K.M. (1996). Levels of evaluation: Kirkpatrick, Kaufman and Keller, and beyond. *Human Resource Development Quarterly*, 7(2), 179-183.

Stufflebeam, D.L. (2001). *Evaluation models*. New Directions for Evaluation, 89. San Francisco: Jossey-Bass.

Tamkin, P., Yarnall, J., & Kerrin, M. (2002). *Kirkpatrick and beyond: a review of models of training evaluation*. Report 392, Institute for Employment Studies, Brighton, UK.

Warr, P., & Bunce, D. (1995). Trainee characteristics and the outcomes of open learning. *Personnel Psychology*, 48(2), 347-375.

Warr, P., Allan, C., & Birdi, K. (1999). Predicting three levels of training outcome. *Journal of Occupational and Organizational Psychology*, 72(3), 351-375.