Reinventare la valutazione

Giampietro Gobo

Uno strumento imprescindibile per il buon esito di un intervento pubblico è dato dalla valutazione ex ante delle sue condizioni di successo, o dalle valutazioni ex post di altri interventi simili nel passato. Ma cosa s'intende con valutazione? È possibile dare una definizione di questo termine che ne circoscriva il significato al di là dei vari approcci esistenti?

In letteratura esistono vari concetti di valutazione, anche molto diversi fra loro. Secondo Mauro Palumbo (uno fra coloro che meglio hanno approfondito il concetto) la valutazione è "un'attività cognitiva tesa a fornire un giudizio su un'azione (o complesso di azioni coordinate) da svolgersi intenzionalmente e destinata a produrre effetti esterni"; essa "usa gli strumenti delle scienze sociali e segue procedure rigorose e codificabili". Mentre il carattere della valutazione varia a seconda dell'approccio adottato (positivista, pragmatista, costruttivista ecc.) e a seconda dell'attore istituzionale che la effettua (o la commissiona), il compito della valutazione (in tutte le sue accezioni) è produrre un giudizio. La valutazione ha finalità orientative (ci dice cosa fare per migliorare) e strumentali (serve a prendere una decisione); questo la distingue sia dalla ricer-

Giampietro Gobo è professore straordinario di Metodologia della ricerca sociale e insegna Valutazione presso l'Università di Milano. Si occupa di metodi quantitativi e qualitativi, con particolare attenzione a raccolta e costruzione dei dati. Su questo tema ha pubblicato *Doing Ethnography* (Sage, 2008) e, con Sergio Mauceri, *Collecting Survey Data* (Sage, 2014).

ca di base (che ha finalità puramente cognitive) sia dalla ricerca applicata (che è funzionale al committente, e non tiene conto di tutti gli stakeholder). La valutazione è anche un'altra cosa sia rispetto al monitoraggio, che è un controllo su tempi e procedure, sia rispetto dall'*audit*, che di solito implica un controllo di conformità (di uno stato di cose) con una norma o direttiva. Si distingue infine dal *benchmarking* (il confronto con le attività di altre organizzazioni), dalla certificazione e dal *rating* (che assegna voti alle prestazioni). Tuttavia, la valutazione può basarsi anche su questi strumenti e tecniche, inglobandoli, essendo un'attività di tipo descrittivo (come il monitoraggio), ma dotata di valore normativo (come l'*audit*) e causale (come la spiegazione scientifica).

Posto che la valutazione combina (e integra) alcune di queste tecniche, perché ritiene che occorra reinventarla?

Le rispondo con un esempio. Comunemente si pensa che un buon corso di laurea dovrebbe perdere pochi studenti per strada. Ne segue che il numero degli iscritti (anno dopo anno) sia un buon indicatore della qualità di un corso: se partono in cento, l'ottimo sarebbe che tutti e cento arrivassero alla laurea. Prendiamo ora il caso (reale) di un corso di laurea in biologia e poniamo che al primo anno s'iscrivano 100 matricole; al secondo anno il corso perde il 30 per cento degli studenti. Questo semplice dato sarebbe, secondo un'opinione comune, un valido indicatore del fatto che il corso in questione ha delle gravi carenze e necessita degli interventi per abbassare la quota degli abbandoni. Tale valutazione sembra ragionevole e del tutto plausibile; tuttavia, è astratta e incapace di cogliere la realtà. Aggiungiamo allora un po' di concretezza, o meglio: aggiungiamo il contesto. Sempre nello stesso ateneo c'è un corso di medicina. Entrare a medicina è assai difficile: bisogna superare un test decisamente selettivo, per cui molti candidati rimangono fuori. I corsi di medicina e di biologia hanno vari insegnamenti in comune, così diversi studenti che non hanno passato il test di medicina si iscrivono a biologia, sostengono tutti gli esami del primo anno e al secondo anno passano a medicina, scavalcando il test d'ingresso (la normativa di Ateneo lo permette). Usciti dalla

porta, rientrano dalla finestra. Conoscendo il contesto, capiamo che il 30 per cento di abbandoni del corso di biologia non ha niente a che vedere con la qualità di questo corso di laurea. Molti di coloro che lo abbandonano sono entrati fin dall'inizio con l'intenzione di passare a medicina, indipendentemente dalla qualità del corso di biologia. Cosa ci insegna questo esempio? Che una valutazione standard come quella attuale (gli stessi parametri di giudizio per tutti i corsi di laurea sul territorio nazionale) non ha senso: anzi, è deleteria e fuorviante. Per questo occorre reinventare la valutazione, ampliare e superare i concetti e i metodi che stanno alla base della valutazione convenzionale, per passare a una valutazione contestuale. Perché i contesti contano.

Cosa comporta, concretamente, la valutazione contestuale?

Noi viviamo sempre in un contesto; agiamo in situazioni particolari e specifiche (mai in contesti generali, un evidente ossimoro). Le nostre azioni sono contesto-dipendenti. Se la valutazione vuole essere accurata e rigorosa, non può che seguire i suoi soggetti là dove agiscono, dove i significati prendono forma. In altre parole, occorre portare la società (le dinamiche sociali e organizzative) nella valutazione. L'esempio appena considerato è illuminante. Una valutazione basata esclusivamente su strumenti standard (come le statistiche) senza l'ausilio dell'etnografia direbbe che il corso di biologia ha gravi carenze. Abbiamo bisogno dell'etnografia (e dei metodi qualitativi) per capire il fenomeno: occorre conoscere il contesto di quell'Ateneo, andare a intervistare un campione di studenti, parlare con i docenti ecc. Tutte cose che la valutazione partecipata in parte fa. È importante precisare che quando parlo di statistiche mi riferisco a due realtà tra loro diverse. Una cosa sono le statistiche sociali che provengono da questionari e da interviste (*survey data*), altra cosa sono le statistiche ufficiali che si ricavano

> *"Se la valutazione vuole essere accurata e rigorosa, non può che seguire i suoi soggetti là dove agiscono, dove i significati prendono forma."*

dai documenti prodotti da organizzazioni e amministrazioni (*process-produced data*), come per esempio il numero di iscritti, dipendenti, segnalazioni, reclami o altro. Nel primo caso il ricercatore si costruisce il dato, se lo va a prendere (per così dire) in modo indipendente dalle fonti istituzionali. Nel secondo caso usa quello che l'organizzazione (nel nostro esempio l'Ateneo) ha costruito. In entrambi i casi (ma soprattutto nel secondo) avremo comunque bisogno d'interpretare i dati per mezzo dell'etnografia, altrimenti rischiamo di prendere delle colossali cantonate.

Cosa si intende con contesto, e quali sono i vantaggi della valutazione contestuale rispetto alla valutazione standard?

Per contesto s'intende l'ambiente in cui avviene un'azione, un'attività, una pratica. Essa è sempre inserita, incastonata (*embedded*) in un contesto storico, sociale, economico, politico e organizzativo. Il fatto che il contesto sia una realtà così ampia (olistica) ne costituisce la forza e insieme la debolezza. La sua forza perché solo in questo modo possiamo comprendere un'azione e valutarla; la sua debolezza perché l'analisi si complica, i tempi per l'indagine si allungano e occorrono strumenti metodologici articolati. Tutti aspetti poco amati dai sostenitori degli approcci dominanti, che vorrebbero far tutto in poco tempo, con poche risorse, e soprattutto con strumenti standardizzati (incapaci per loro natura di cogliere i contesti). Uno dei testi adottati nel mio corso è il libro *Da Chernobyl a Linate* di Maurizio Catino, un sociologo dell'organizzazione che prende in esame diversi casi di incidenti (ferroviari, aerei, navali). Egli mostra che essi sono solo in apparenza dovuti a errori umani o black-out tecnologici, essendo tipicamente errori organizzativi. Tuttavia, solo analizzando il contesto organizzativo possiamo capire le cause di questi incidenti e possibilmente rimuoverle affinché non si ripetano. Prendiamo il naufragio della Costa Concordia nel gennaio 2012 davanti all'isola del Giglio (trenta morti e due dispersi). A prima vista, l'incidente sembra interamente dovuto al cattivo operato del capitano della nave, il famoso Schettino. Tuttavia, come ha messo in luce Catino, questa lettura è semplicistica. I fattori causali erano già noti e ricorrenti.

Per esempio, non era la prima volta che una nave della Costa faceva il cosiddetto inchino davanti al Giglio (o in altre località). In passato il comandante della capitaneria di porto, Falco (divenuto famoso per la telefonata in cui intima a Schettino di tornare a bordo della nave), aveva visto la Concordia fare questo inchino e non aveva mai censurato né denunciato il fatto alle autorità. Inoltre, pare che l'azienda Costa Crociere promettesse questo fuori programma ai suoi passeggeri. Diversi elementi dicono, insomma, che l'incidente era latente, e che prima o poi sarebbe successo. Poi, quando succede, si dà la colpa al capitano, e si parla di errore umano, mentre sarebbe più corretto parlare di errore o incidente organizzativo, perché la responsabilità va imputata alla Costa Crociere e alla capitaneria di porto, non a un singolo attore (il cosiddetto fattore umano). Ecco perché propongo un libro di sociologia dell'organizzazione in un corso di valutazione. Serve a mostrare che una valutazione circoscritta soltanto al suo obiettivo o target riconosciuto (gli abbandoni scolastici, la performance dei dipendenti, gli effetti di una legge, i risultati di una politica ecc.), senza una visione globale, olistica, è molto deficitaria.

Posto che il contesto va rimesso al centro, ci sono altri fattori (dietro il successo o l'insuccesso di un intervento) che possono contribuire a spiegarne l'impatto? Quali sono i principali approcci alla valutazione?

Esistono almeno tre approcci alla valutazione: l'approccio positivista-sperimentale (che si focalizza sugli obiettivi di un intervento, di una politica, di un piano), quello pragmatista (che si focalizza sugli standard di qualità) e quello costruttivista (più orientato a comprendere il processo sociale che accompagna e sostanzia un intervento da valutare). Facciamo un esempio per capire le loro differenze. Supponiamo che il sindaco di una grande città voglia realizzare un programma di lotta alla disoccupazione. Insieme all'assessore delle politiche sociali elabora un piano d'intervento: dare per un anno un sussidio di 500 euro mensili ai disoccupati. Passato l'anno, il sindaco desidera valutare l'efficacia dell'intervento. Per fare questa valutazione contatta preventivamente (cioè prima che parta l'inter-

vento) tre esperti di valutazione, uno per ogni approccio. Il primo consulente (che appartiene all'approccio positivista-sperimentale) potrebbe proporre il seguente progetto valutativo. Si prendano due gruppi omogenei di disoccupati. Il primo gruppo (detto "target") viene sottoposto allo stimolo (cioè, riceve il sussidio). Il secondo gruppo (detto "di controllo") non riceve nulla. Passato un anno, e terminato l'esperimento, si analizza la situazione (e il comportamento) dei due gruppi. Se il programma ha avuto successo, il gruppo target dovrebbe esibire un tasso di occupazione superiore al gruppo di controllo. Se invece l'occupazione fra i due gruppi è pressoché uguale, allora vuol dire che il programma ha fallito, che si sono buttati via dei soldi.

Oltre alla totale assenza di una visione contestuale, quali sono i limiti di questo modo di procedere?

In primo luogo, i due gruppi non sono mai omogenei. Quello che possono avere in comune sono alcune caratteristiche (l'età, l'essere disoccupati, il genere, la scolarità), ma molte altre (la capacità d'iniziativa, le reti di conoscenza o capitale sociale, le qualità personali, lo spirito di adattamento, l'inclinazione al sacrificio, la flessibilità mentale) possono essere presenti in modo diversificato e incontrollato nei due gruppi. In secondo luogo, un simile approccio non tiene conto delle variabili intervenienti, nel senso che l'intervento del sindaco potrebbe prendere avvio in una situazione avversa (la chiusura di imprese dovuta al *credit crunch*, alla delocalizzazione degli impianti o al fatto che lo stato non salda i debiti con le aziende) oppure favorevole (una congiuntura internazionale positiva, la nascita di nuove aziende sul territorio): sono variabili che non dipendono dal sindaco e che egli non può controllare. Così, l'occupazione nel gruppo target potrebbe salire o essere comparativamente irrilevante per cause indipendenti dalla bontà del programma. E non saremmo in grado di capire se l'intervento sia stato realmente efficace (a meno di trovare marcate differenze fra i due gruppi). In terzo luogo, esiste un'altra variabile interveniente che potrebbe far fallire il piano, ed è il contesto amministrativo del comune. I sussidi potrebbero essere stati erogati in ritardo o in modo non regolare

per tutto l'anno; l'amministrazione potrebbe richiedere ai disoccupati certificazioni, documenti, attestati o altro che essi non sono in grado di reperire o presentare per tempo. Spesso è la burocrazia a far fallire gli interventi, le innovazioni, le riforme. E focalizzarsi solo sull'intervento senza tenere conto del contesto organizzativo è limitante. In quarto luogo, una simile sperimentazione è troppo rigida, perché se durante l'anno sopraggiungono eventi inaspettati (congiuntura, nuove condizioni, aspetti problematici del piano) non si può far nulla per affrontarli in itinere: si può solo aspettare che la sperimentazione finisca (anche se a metà dell'opera già sappiamo che fallirà) per fare una valutazione ex post. Esiste, infine, un problema morale: come sceglieremo le persone da inserire nei due gruppi? Chi se la sentirebbe di dare 500 euro a Tizio e niente a Caio, solo perché fanno parte di un esperimento?

Passiamo al secondo esperto, il fautore dell'approccio pragmatista.

Il secondo esperto, ancor prima che il programma venga ideato dal sindaco e dalla sua giunta, ne valuterebbe il *worth* e il *merit*. Si chiederebbe, in altre parole, se il piano sia appropriato al contesto (*merit*) e se ne valga la pena (*worth*). Inoltre, analizzando il contesto territoriale, si chiederebbe se ci sono le condizioni (sociali, economiche, politiche) perché il programma possa avere successo e quali siano gli standard di qualità a cui il piano dovrebbe aderire. Questo approccio si definisce pragmatico proprio perché lascia da parte la teoria (le buone intenzioni del sindaco, la bontà intrinseca del piano) e va dritto alla sostanza: l'esistenza o meno delle condizioni di successo di un intervento e la presenza di standard minimi accettabili per la sua realizzazione.

Come procede, infine, il terzo esperto, quello costruttivista?

L'approccio costruttivista si focalizza sul processo sociale, stabilendo che il successo di un intervento deve essere definito da tutti i partecipanti a esso, i cosiddetti stakeholder (cioè portatori di un qualche interesse nel piano). Occorre quindi raccogliere il parere

non solo del sindaco, ma anche dell'amministrazione, dei disoccupati ecc. In questo modo il piano di intervento viene condiviso con i disoccupati (che invece nel primo caso non sapevano neanche di partecipare a un esperimento). Essi vengono accompagnati durante tutto l'anno di sperimentazione. Mediante incontri periodici con tutti gli stakeholder si prendono in considerazione le varie problematiche che sorgono man mano che il piano prosegue attivando dei correttivi. In questa prospettiva il piano non è qualcosa di esterno agli stakeholder (non è qualcosa calato su attori passivi) ma è costruito processualmente anche da loro, che sono quindi parte attiva del piano, contribuiscono a costruirlo, possono farlo fallire o funzionare. Ecco perché è necessario alimentare un continuo dialogo tra i vari stakeholder, con l'aiuto del valutatore in qualità di facilitatore.

Sembra un approccio complesso, non facile da realizzare. Quali sono i suoi principi di fondo?

Diciamo che è un approccio non ingenuo nei confronti della realtà. E si basa su quattro elementi. In primo luogo, la realtà non è indipendente dal valutatore (osservatore): questi è (anche se non lo sa) parte attiva, soggetto interno e non esterno al piano che dovrebbe valutare. I suoi metodi di valutazione (anche se inconsapevolmente) costruiscono i risultati stessi. Usando un questionario otterrà risultati diversi da quelli che otterrebbe facendo un focus group o un'etnografia. In secondo luogo, ne consegue che la valutazione è un'attività di co-costruzione tra valutatore e stakeholder. In terzo luogo, l'oggetto stesso della valutazione viene costruito processualmente fra tutti gli stakeholder (valutatore compreso). Nell'approccio costruttivista a valutare un programma sono anche i destinatari del programma stesso (dagli utenti agli stessi dipendenti), non solo il valutatore. I risultati della valutazione si creano processualmente nell'ambito delle interazioni tra ricercatore e partecipanti. Il ricercatore quindi sente tutte le parti

> *"La valutazione è un'attività di co-costruzione tra valutatore e stakeholder."*

(sia i destinatari che gli esecutori del programma), ne ascolta le opinioni e diventa negoziatore, facilitatore, mediatore delle diverse opinioni e prospettive. Gli obiettivi della valutazione sono da una parte l'apprendimento: comprendere la situazione, interpretare il pensiero degli attori e imparare qualcosa su come funzionano certi piani di intervento (problemi, lacune, difficoltà, inadeguatezze) per migliorarli in futuro; dall'altra il potenziamento (*empowerment*) delle capacità dei beneficiari del piano. In quarto luogo, la valutazione è primariamente un problema teorico e solo secondariamente un problema tecnico, di strumenti.

Cosa significa, concretamente, trattare la valutazione come un problema teorico, prima ancora che tecnico?

Come in passato hanno sostenuto diversi autori, fra cui Pawson, Tilley, Chen, Rossi e Weiss, significa riconoscere che la valutazione è sempre un'azione orientata da una teoria (*theory-driven* o *theory based*), purtroppo spesso non esplicitata, ma non per questo ininfluente. Porsi il problema teorico significa cercare di esplicitare l'idea sottostante per cui il piano dovrebbe funzionare, i motivi per cui l'esperimento (o il quasi-esperimento) con i due gruppi di disoccupati dovrebbe aver successo. Si tratta quindi eminentemente di un problema di concettualizzazione. Magari, nel mondo reale, gran parte di coloro che ricevono un incentivo non lo cercano nemmeno un lavoro; magari gli altri lo cercano, senza incentivi, e lo trovano. Molto spesso veniamo a sapere che certe politiche pensate per risolvere un problema non funzionano: forse era inadeguata l'idea sottostante (di solito implicita). Questo accade perché i contesti reali non vengono quasi mai presi in considerazione, oppure vengono valutati in maniera astratta. Il problema ha cominciato a emergere agli inizi degli anni Novanta, quando alcuni autori (non tutti costruttivisti) hanno cominciato a criticare l'approccio positivista-sperimentale e a sottolineare che quello che si doveva prioritariamente valutare erano gli assunti (teorici) sottostanti al piano d'intervento, non solo i risultati finali dello stesso. Pawson e Tilley, autori nel 1997 del famoso *Realistic Evaluation*, sono gli esponenti di punta di quel filone teorico.

Può fare un esempio di teoria sottostante, o di idea alla base di un piano d'intervento?

Se voglio valutare la didattica universitaria devo avere un'idea di cosa sia una buona e una cattiva didattica. La mia idea di cosa sia una buona didattica (e tale idea è spesso tacita, non sempre trasparente e consapevole nella testa del valutatore) guiderà la mia concettualizzazione dell'idea di didattica, la sua operazionalizzazione in indicatori e variabili ecc. Lo stesso dicasi per la valutazione del merito: come faccio a valutare il merito se non definisco in precedenza il concetto di merito? E questa è chiaramente una questione teorica (e in ultima analisi politica, in senso lato). Ciò vuol dire che la valutazione dovrebbe anzitutto chiarire la "teoria sociale" (Chen) che sta dietro il programma d'intervento, con i meccanismi sociali su cui si fonda e che dovrebbero portare il programma al raggiungimento dei risultati, e poi controllare la teoria stessa (non il programma) tramite uno studio.

Che rapporto ha il contesto sociale con tutto questo?

Prendiamo un esempio, diventato un classico della valutazione realistica: l'efficacia delle telecamere a circuito chiuso (CCTV) nel combattere la criminalità. Decisori e (molti) valutatori, condividono un atteggiamento positivista che ritiene che esistano relazioni di causa-effetto lineari tra una determinata azione e il successivo cambiamento sociale. Tuttavia, non è raro che il cambiamento avvenuto (B) sia l'effetto non tanto dell'azione (A) in sé, quanto dell'intervento di altri meccanismi sociali (o *intervening variables*). Poniamo il caso di un parcheggio in cui vengono installate delle CCTV e successivamente si nota una riduzione dei furti. A prima vista, si potrebbe pensare che l'introduzione delle CCTV (A) abbia prodotto il calo dei furti (B). Tuttavia, le telecamere in sé non hanno questo potere taumaturgico. Come scrivono Pawson e Tilley, "non c'è nulla nell'installazione di una CCTV all'interno di un parcheggio che di per sé possa scoraggiare il crimine". È necessaria la concomitanza di particolari condizioni affinché le telecamere siano realmente efficaci. Primo, bisognerà che ci sia qualcuno (un poliziotto, un addetto alla sicurezza) che in tempo

reale osservi le immagini trasmesse dalle telecamere e intervenga subito, altrimenti un ladro che indossi un passamontagna (o un berretto con visiera) ben calzato potrà agire indisturbato e senza rischiare di essere riconosciuto. Secondo, bisognerà che le telecamere siano collegate con una stazione di polizia situata nelle immediate vicinanze, in modo che un agente possa intervenire. Terzo, bisognerà che il parcheggio sia privo di ostacoli visivi per le telecamere (colonne, guardrail) dietro i quali il ladro potrebbe agire indisturbato. Quarto, bisognerà che ci sia nel parcheggio un'adeguata illuminazione, altrimenti le riprese non saranno abbastanza nitide e sarà difficile riconoscere il ladro. Quinto, bisognerà che i ladri siano persone del luogo; se vengono da fuori (magari da molto lontano), potrebbero infischiarsene di essere ripresi: i loro volti solitamente non sono presenti negli schedari della polizia locale (che ha le foto dei criminali locali, oltre a quelle di terroristi o serial killer), quindi il rischio di essere catturati è basso. Ecco, in assenza di queste condizioni, le telecamere possono fare (e spiegare) ben poco. Il calo dei furti riscontrato potrebbe essere dovuto (invece, o anche) all'intervento di varie concause. In altre parole, la presenza delle telecamere potrebbe aver agito in modo indiretto: o spingendo i proprietari a ricordarsi di chiudere l'auto (alcune assicurazioni non risarciscono il cliente disattento, e la sua dimenticanza potrebbe essere immortalata dalle telecamere), o aumentando il passaggio di clienti nel posteggio, accrescendo così il controllo sociale (se si sparge la voce che nel parcheggio ci sono le telecamere, e la gente lo reputa un luogo più sicuro, ci sarà maggior presenza di clienti e quindi maggiore controllo sociale: sarà più difficile per i ladri agire indisturbati).

Come si traduce questa molteplicità di considerazioni quando passiamo a definire un framework generale di valutazione?

Pawson e Tilley sostengono che il successo o il fallimento di un'iniziativa, ovvero il suo risultato (*outcome*), è una funzione di due fattori combinati: un meccanismo (M) sottostante al programma e il contesto (C) nel quale il programma viene applicato. La loro

formula è O = f(M,C). Con meccanismo sottostante (M) bisogna intendere le risorse (materiali, cognitive o emotive) che l'iniziativa fornisce nella convinzione che esse possano influenzare le azioni dei soggetti che (nei vari ruoli) prenderanno parte al programma. Il funzionamento (o il mancato funzionamento) di questo meccanismo dipende, come detto, dal contesto (C), cioè dalle caratteristiche tanto dei soggetti quanto del luogo in cui si svolge il programma. Questo modello di valutazione permette ai ricercatori di raccogliere informazioni su tre aspetti cruciali. Primo, che cos'è che funziona (in un programma); secondo, per chi (per che tipo di persone) funziona; terzo, in quale contesto funziona. In altri termini, affinché un programma o un'innovazione abbiano successo, occorre che si attivino certe risorse (il meccanismo che deve supportare l'innovazione) e contemporaneamente che si realizzino certe condizioni (il contesto sociale, politico e amministrativo). Pawson spiega infatti che un programma, considerato in sé, può fare ben poco: sono i meccanismi dei programmi a innescare il cambiamento, non i programmi in quanto tali.

Un problema comune a tutte le teorie della valutazione è la definizione di corrette unità di misura. Per esempio, nel dibattito sulla valutazione della ricerca scientifica si assiste da tempo a scontri molto accesi sulla fondatezza metodologica di indici e parametri bibliometrici usati per valutare la produttività dei ricercatori. Da metodologo, come giudica l'utilizzo di questi parametri?

Gli indicatori e gli indici usati nella valutazione della ricerca sono costruzioni molto strane, anche dal punto di vista matematico. Negli ultimi anni, i parametri che si sono imposti nella comunità scientifica (non senza critiche, talora accese) sono due: uno è il cosiddetto *impact factor* e riguarda la valutazione delle riviste (o collane) su cui pubblicano ricercatori e scienziati. L'altro si chiama "H-index" (o indice di Hirsch): è stato ideato nel 2005 dall'argentino Jorge E. Hirsch (un fisico che insegna alla California University di San Diego) e concerne la valutazione della produzione scientifica di un individuo. In entrambi i casi, più questi indici sono

numericamente alti, più prestigio viene attribuito alla rivista o all'individuo valutati. Il primo è stato oggetto, nei decenni, di così tante critiche che in Gran Bretagna si è deciso di rinunciare del tutto a usare gli *impact factor* e le classifiche delle riviste nel nuovo esercizio di valutazione della ricerca (il REF del 2014). Il secondo indice ha avuto molto successo nel campo della medicina, della chimica, della fisica, della biologia e dell'economia, tutte discipline che usano la matematica senza però aver sviluppato una riflessione sulla matematica, e che soffrono di una scarsa consapevolezza metodologica su come si costruisce il dato. Le critiche più forti contro questi parametri bibliometrici vengono proprio dai matematici. In Italia, per esempio, il matematico Giorgio Israel sostiene che la matematica è utile per analizzare le proprietà di oggetti inanimati, ma inutile per comprendere gli esseri umani. Egli pensa che sia poco fondato (e addirittura fuorviante) valutare la qualità della ricerca mediante proprietà quantitative. Anche l'indice di Hirsch ha ricevuto molte critiche e prevedo che in un futuro non lontano possa fare la fine dell'*impact factor*, venendo lentamente abbandonato. Accanto alle molte critiche puntuali e circoscritte che ha ricevuto (per una sintesi si veda il libro di Baccini, *Valutare la ricerca scientifica*) ne voglio evidenziare una che mi pare non sia stata adeguatamente evidenziata e che riguarda i suoi assunti (culturali) taciti: l'indice H guarda maggiormente alla quantità invece che alla qualità della produzione scientifica (che si potrebbe derivare "semplicemente" dalla lettura del contenuto di una pubblicazione). Esso premia infatti coloro che scrivono molto e vengono letti o citati da poche persone. Sembra un paradosso, eppure succede proprio questo. L'algoritmo sottostante è sconosciuto e fa anche molti pasticci. Per farla breve, un indice H 10 vuol dire che 10 tue pubblicazioni sono state citate almeno 10 volte (la difficoltà di spiegazione, per me, e di comprensione, per il lettore, mostra la perversione di questa logica). Ebbene, la cosa davvero singolare è che, per l'indice H, è

"Le critiche più forti contro questi parametri bibliometrici vengono proprio dai matematici."

pressappoco la stessa cosa che una pubblicazione sia citata 10 volte oppure 5000, perché entrambi i risultati fanno crescere l'indice sempre e solo di un punto. Se non mi crede, faccia la prova con Tuhiwai Linda Smith, un'autrice che ha pubblicato (fra le altre cose) questi due testi: un libro intitolato "Decolonizing methodologies: Research and indigenous peoples", con 5924 citazioni al 13/09/2013, e un articolo intitolato "New mythologies in Maori education", con 14 citazioni al 13/09/2013. Scoprirà che le citazioni ricevute dalla prima pubblicazione non contano, e questo è assurdo. Per quanto mi riguarda, sia dal punto di vista accademico-scientifico che dal punto di vista della reputazione, preferirei di gran lunga avere una sola pubblicazione citata migliaia di volte (che mi attribuisce un indice H pari a 1) che una decina di pubblicazioni citate una decina di volte (che rende il mio indice H pari a 10). E se dovessi assumere un ricercatore, non avrei dubbi: farei l'esatto opposto di quello che suggerisce l'indice H, che secondo me favorisce, stimola, incrementa la mera quantità a scapito della qualità, e in definitiva promuove la mediocrità.

Ha detto che occorre portare la società (ovvero le dinamiche sociali e organizzative) dentro il processo di valutazione. Cosa ha a che fare la società con la valutazione accademica?

C'entra molto da vicino. E ad almeno due livelli: il primo teorico (gli assunti culturali taciti, sottostanti a qualsiasi indice o criterio di valutazione); l'altro contestuale (come funzionano e agiscono nella società questi criteri e indici). Secondo il senso comune accademico, un buon ricercatore è qualcuno che possiede (almeno) le seguenti caratteristiche: fa tanta (e buona) didattica; pubblica molti contributi (si spera di qualità e su riviste prestigiose); svolge buona ricerca; vince bandi nazionali e internazionali; partecipa a convegni nazionali e internazionali; svolge anche un lavoro istituzionale; partecipa alla vita del dipartimento e della facoltà (seminari, convegni ecc.). A prima vista, sembrano criteri ragionevoli. Tuttavia, se cominciamo a sviscerare gli assunti culturali (taciti o espliciti), ovvero il modello mentale sottostante a questo particolare concetto

di buon ricercatore, ci possiamo chiedere: chi meglio potrebbe performare (a parità d'intelligenza) su tutte queste aree? In un ranking ragionevole e sociologicamente sensato, penso che arriverebbe senz'altro primo una persona single; secondo una persona con partner, senza figli; terzo una persona con partner, con 1 figlio; quarto una persona con partner, con 2 figli; e così via. Tralasciamo, ovviamente, tutte altre possibili combinazioni (avere un famigliare malato o disabile, un partner lontano per lavoro, il supporto dei genitori, condizione economiche agiate che permettono baby-sitter a gogò, ecc.), perché altrimenti otteniamo una complessità di situazioni difficilmente trattabile con gli strumenti attuali. Ora, il modello culturale sottostante (tacito), incorporato nei criteri valutativi sopra menzionati, incarna un profilo di ricercatore molto più simile a una Rita Levi Montalcini (categoria 1) o Margherita Hack (categoria 2), che a una persona con 3 figli. Tuttavia, è ragionevole ipotizzare che chi ha figli, a parità di capacità intellettuali, abbia avuto un rallentamento nella produzione scientifica (senza pensare anche a una sua ridotta capacità di ricerca e presenza istituzionale)? Decine di ricerche documentano questo effetto, e in particolare che le donne hanno minori pubblicazioni degli uomini. Questo perché la loro condizione, in una società ancora profondamente asimmetrica nella distribuzione delle responsabilità famigliari e nelle pratiche di cura, è socialmente penalizzante. Un discorso che, ovviamente, non vale solo per l'accademia, ma per il mondo professionale nel suo complesso.

Tornando al rapporto fra valutazione e contesto sociale, che relazione c'è fra valutazione contestuale e valutazione partecipata?

Sono due cose distinte analiticamente, ma interrelate. La valutazione contestuale si pone l'obiettivo di ampliare lo sguardo, conglobando il contesto, ma potrebbe essere fatta anche da un solo ricercatore. La valutazione partecipata, invece, ha come obiettivo il coinvolgimento di tutti gli stakeholder nel processo valutativo. Per questa ragione, sono due approcci che ben si compenetrano e si arricchiscono reciprocamente. Hanno in comune un punto negati-

vo: non basiamo tutto sulle statistiche ufficiali. I dati delle statistiche ufficiali vanno contestualizzati, e per farlo adeguatamente abbiamo bisogno del contributo degli attori sociali, che sono immersi nelle realtà che si tratta di valutare. Ripeto, il punto non è fare a meno delle statistiche ufficiali, ma costruirle etnograficamente, farle interagire con le inchieste (o sondaggi) e, soprattutto, con la ricerca qualitativa. La differenza fra inchieste (o sondaggi) e statistiche ufficiali sta nel fatto che in un sondaggio (soprattutto se discorsivo) possiamo almeno dire che i soggetti hanno parlato; magari hanno frainteso le domande, però almeno hanno potuto esprimersi. Nel caso delle statistiche ufficiali non abbiamo neanche incontrato le persone. Purtroppo, il vero limite dei sondaggi tradizionali è che spesso si preparano i questionari senza prima aver fatto una ricerca qualitativa (ecco perché la tecnica del sondaggio discorsivo, che ho proposto qualche anno fa, è una valida alternativa). La funzione della ricerca qualitativa è quindi duplice: può precedere il sondaggio (consentendoci di costruire un questionario adeguato) e può aiutarci (ex post) a interpretarne i risultati.

“Il punto non è fare a meno delle statistiche ufficiali, ma costruirle etnograficamente, farle interagire con le inchieste (o sondaggi) e, soprattutto, con la ricerca qualitativa.”

Forse per questo una lunga tradizione preferisce combinare una serie di metodi diversi, per evitare le distorsioni dovute ai limiti di un singolo metodo? È questa la forza dei metodi misti?

Direi di sì. Non dimentichiamoci che la survey europea aveva in origine un’impostazione *multi-method*. A partire dal 1850, in Francia e Inghilterra sono nati gli studi sulla povertà e la condizione operaia. Si facevano rudimentali questionari, ma la cosa importante era osservare i poveri nei pub, nelle case, nei centri di ritrovo, nelle chiese, a scuola (contando in ogni luogo gli uomini, le donne, i bambini); sono queste le statistiche etnografiche. Oppure si chiede-

va alle persone di fare dei budget: calcolare quanto avevano speso nell'ultima settimana. Questo per dire che la survey europea nasce fin dall'inizio come metodo misto. Poi fra il 1930 e il 1950 è arrivata dagli Stati Uniti una survey monometodo. E soltanto a partire dal 1990 è rinato l'interesse per i *mixed methods*, che oggi sembrano vivere una nuova stagione.

Sarebbe possibile, oggi, realizzare questo modello di valutazione partecipata in una grande organizzazione, quale, per esempio, una multinazionale o la stessa Pubblica amministrazione?

È quello che viene fatto, con successo, da più di vent'anni. Negli anni Novanta del secolo scorso, lo psicologo sociale Edgar Schein ha introdotto il concetto di "consulenza di processo", che si differenzia dall'approccio tradizionale alla consulenza (l'esperto esterno che viene a insegnare come si fanno le cose, dove bisogna cambiare). Il consulente di processo ha il ruolo maieutico di tirar fuori, far circolare, mettere in relazione, sbloccare le conoscenze che già ci sono ma per tante ragioni non emergono. Quindi raccoglie informazioni in loco, parla con il personale, ascolta le loro storie, li fa discutere. Invita il management a organizzare gruppi di discussione, per confrontarsi con i dipendenti, facendo in modo che tutti possano parlare, e che anche il personale delle pulizie possa dire la sua (garantendo che ciò che dice avrà pari dignità rispetto alle idee del dirigente); perché anche l'usciere può avere delle buone idee. Il punto non è imporre dall'esterno delle soluzioni già pronte (che sono spesso decontestualizzate), ma far emergere le connessioni e i problemi nascosti. La consulenza di processo è molto in sintonia con l'approccio costruttivista, un approccio olistico adatto ad affrontare realtà e problemi complessi. Esso analizza il contributo

> *"Il consulente di processo ha il ruolo maieutico di tirar fuori, far circolare, mettere in relazione, sbloccare le conoscenze che già ci sono ma per tante ragioni non emergono."*

di tutti gli attori coinvolti nel programma (gli stakeholder) e osserva come si trasforma (cosa diventa) un programma quando viene applicato. La valutazione è vista come un processo da cui possiamo imparare a fare meglio, il cui scopo è l'apprendimento, la crescita della conoscenza sul funzionamento di un programma. Siccome quando si attua un programma esso muta al contatto con il contesto (producendo conseguenze inattese), non avrebbe alcun senso valutarlo con lo sguardo iniziale. I criteri di successo di un programma devono essere riferiti alla situazione specifica e trovati ex post (non definiti ex ante).

La valutazione fa (ancora) paura?

La valutazione spaventa per tanti motivi; alcuni sani, altri meno. La valutazione fa paura a chi ha rendite di posizione, a chi preferisce lo status quo, a chi non vuol cambiare. Tuttavia, fa paura anche alle persone oneste, a chi fa seriamente il proprio lavoro, perché viene presentata come una sorta di castigamatti, secondo lo stereotipo positivista. L'approccio costruttivista delinea invece un tipo di valutazione cooperativa, che gli attori sociali possono fare solo insieme ai valutatori. Se la valutazione viene percepita dai valutandi come una sorta di giuria che emette condanne, essi faranno di tutto per difendersi e nascondere i problemi. È questa la conseguenza generale della *blame culture* di cui parla Catino, la cultura della colpa individuale. A tal proposito, Catino porta un esempio illuminante. Fino a non molto tempo fa nella nostra Aeronautica Militare si faceva fatica a introdurre una cultura della sicurezza tesa a rilevare tutti i possibili incidenti. Finché vigeva una logica della colpa individuale, ognuno cercava di occultare gli incidenti scampati, gli errori personali ecc. per paura di essere sanzionato. Quando si passò a una *no-blame culture*, le segnalazioni di anomalie crebbero esponenzialmente. Qualcuno potrebbe dunque pensare che l'effetto sia stato negativo, che prima c'erano meno segnalazioni e quindi il sistema in precedenza era più efficiente. In realtà, i problemi c'erano anche prima, solo che venivano nascosti. Questo principio relativo alla prevenzione degli incidenti organizzativi può essere applicato alla valutazione in generale. Nella valutazione cooperativa e

partecipata, valutatore e valutando possono (e dovrebbero) scegliere insieme il modello valutativo (di performance, produttività o altro) e gli indicatori. Il valutando conosce il contesto molto meglio del valutatore e sa quali strumenti sono più adatti per una corretta valutazione. Il valutatore, dal canto suo, deve fare attenzione a non imporre un suo schema valutativo: dovrebbe cercare, di volta in volta, di costruirne uno insieme al valutando. Nella mia esperienza, quando si segue questo approccio, i dipendenti delle organizzazioni (pubbliche o private) tendono a rispondere positivamente.

Testi consigliati

Baccini A., *Valutare la ricerca scientifica. Uso e abuso degli strumenti bibliometrici*, Bologna, Il Mulino, 2010.

Catino M., *Da Chernobyl a Linate. Incidenti tecnologici o errori organizzativi?*, Milano, Bruno Mondadori, 2008.

Gobo G., Ritorno a Likert. Verso una *survey* discorsiva, *Sociologia e Ricerca Sociale*, XXX, n. 88, 2009, pp. 5-28.

Palumbo M., *Il processo di valutazione. Decidere, programmare, valutare*, Milano, Franco Angeli, 2001.

Pawson R., Tilley N., *Realistic Evaluation*, London, Sage, 1997.

Schein E.H., *La consulenza di processo. Come costruire le relazioni d'aiuto e promuovere lo sviluppo organizzativo*, Milano, Cortina, 2001.

Reinventare_Gobo.indd 60 17/10/2013 16:34:59