

SOMMARIO

Mita Marra: *“I confini porosi del dibattito inclusivo sui metodi valutativi”*

Alberto Martini: *“Bicchiere mezzo pieno o mezzo vuoto? Stiamo almeno parlando dello stesso bicchiere?”*



I confini porosi del dibattito inclusivo sui metodi valutativi Idee per una cultura della valutazione

Mita Marra - Presidente dell'Associazione Italiana di Valutazione

Come è noto il dibattito sui metodi e sulle tecniche di valutazione nella comunità internazionale dei valutatori è sempre particolarmente acceso e talora polarizzato tra coloro che si schierano tra i randomistas e coloro che si iscrivono nell'ambito degli approcci basati sulla teoria del programma. I randomistas sono i fautori dell'approccio sperimentale, che tra i metodi della ricerca sociale è quello più robusto, in grado di risolvere il problema dell'attribuzione. Secondo tale approccio, infatti, per attribuire gli effetti osservati alla causa applicata — un programma o una politica pubblica— occorre distribuire casualmente i beneficiari in gruppi di trattamento e di controllo e verificare le differenze. Diversamente, i valutatori che ricostruiscono la teoria del programma partono dal presupposto che i programmi perseguono strategie complesse la cui razionalità non si riduce nell'immediata valutazione dei costi e benefici sociali. I beneficiari dei programmi e gli

attori che prendono parte alle arene pubbliche rispondono alle opportunità che le politiche offrono, godendo di una considerevole autonomia che si può tradurre in notevoli differenze di attuazione dei programmi. Diventa di fondamentale importanza ricostruire, allora, i principi e le motivazioni che guidano il loro agire. Occorre interpretare le percezioni e le motivazioni che contribuiscono a spiegare perché una politica funziona in un contesto e non funziona in un altro.

In realtà, il dibattito sui metodi e sugli approcci valutativi è molto più segmentato e inclusivo di quanto non si pensi. Esiste un'elevata eterogeneità nella comunità dei valutatori che possono scegliere i metodi — analisi statistiche o qualitative, di rete, organizzativi e di caso, femministi, di genere e sensibili alle differenze culturali, nonché approcci sistemici in linea con il paradigma della complessità — in relazione agli obiettivi conoscitivi.

Benché esistano dei “confini” che distinguono i modelli concettuali, le competenze, le pratiche e gli ambiti di applicazione dei vari metodi e approcci valutativi, tali confini diventano sempre più porosi e sfumati nelle esperienze concrete di valutazione, particolarmente in quelle più innovative e collaborative. I confini sono salienti per apprendere a padroneggiare tecniche e metodi avanzati, per orientarsi nel decifrare i molteplici risvolti cognitivi dei programmi e dei contesti, ma ad un certo punto, per evitare di reificarli, occorre andare al di là degli steccati disciplinari, delle rigide classificazioni metodologiche, delle etichette retoricamente utili alle tecnocrazie... Bisogna proprio intenzionalmente “violare” i confini, come ha sempre suggerito Hirschman a proposito dell'arte del trespassing.

Questo numero 10 della newsletter AIV presenta il solo contributo di Alberto Martini, che ritornando sul controverso dibattito sulla valutazione dei fondi strutturali,

propone le sue riflessioni sull'approccio controfattuale. Le considerazioni avanzate — con cui si è liberi di concordare o meno — sembrano costruire ponti metodologico-concettuali per affrontare in maniera collaborativa problemi valutativi sempre più complessi e sfidanti.

Il prossimo numero 11 della Newsletter AIV ripropone i contributi presentati al seminario sul tema delle nuove competenze del Senato della Repubblica e preannuncia una specifica sessione del prossimo Congresso dedicata all'approfondimento del tema delle riforme costituzionali.

Invitiamo i lettori ad intervenire con commenti agli articoli e a suggerire approfondimenti ai temi trattati.

Per proporre contributi scrivete a:

segreteria@valutazioneitaliana.it

Bicchiere mezzo pieno o mezzo vuoto? Stiamo almeno parlando dello stesso bicchiere?

Alberto Martini* - Università del Piemonte Orientale "Amedeo Avogadro" (alberto.martini56@gmail.com)

La Rassegna Italiana di Valutazione (RIV), la rivista ufficiale dell'Associazione Italiana di Valutazione (AIV), apre il 2015 dedicando un numero monografico (il n. 58) all'analisi controfattuale. Questo fatto è di per sé un segnale positivo: anche solo cinque anni fa un'iniziativa del genere sarebbe stata impensabile.

Ma un segnale di cosa, esattamente? Due le ipotesi in campo: un interesse genuino per quello che questo approccio potrebbe offrire, oppure solo la rassegnata constatazione della sua

crescente diffusione anche in Europa, incluse strutture importanti della Commissione Europea. Pensiamo alla *DG Employment*, che gestisce il Fondo Sociale Europeo e ha di recente aperto un centro studi dedicato ad approfondire questo tema (il CRIE, Center for Research on Impact Evaluation) o la *DG-Regio* che gestisce il Fondo Europeo per lo Sviluppo Regionale (FESR).

La domanda è tutt'altro che retorica, quindi, viste le reazioni dell'AIV alle critiche impietose di Roberto Perotti e Filippo Teoldi contenute in un articolo apparso su *lavoce.info* nel luglio del 2014.

In estrema sintesi, Perotti e Teoldi sostengono che le valutazioni dei fondi strutturali condotte in questi anni hanno prodotto tanti numeri che non dicono nulla, e in particolare non dicono nulla sugli effetti prodotti dall'utilizzo dei fondi. Tali valutazioni lasciano quindi aperta, o meglio spalancata, la porta alla possibilità che i fondi strutturali non ne abbiano proprio nessuno, di effetto.

Per tentare di richiudere quella porta sarebbe stato necessario investire in rigorose analisi di tipo controfattuale, ma ciò non è stato fatto e dunque il dubbio di una inefficacia diffusa dei fondi rimane. Per questa e altre affermazioni i due autori sono stati tacciati di superficialità e di pressapochismo.

Perotti e Teoldi sono capacissimi di difendersi da soli, quindi non farò loro da avvocato difensore. Ritengo tuttavia sia utile partire da questo episodio perché le critiche mosse a caldo a Perotti e Teoldi mettono chiaramente in rilievo due diverse motivazioni per l'avversione all'approccio controfattuale, che occorre tenere distinte: una di tipo metodologico e l'altra di tipo, diciamo così, professional-corporativo.

La prima motivazione è di carattere epistemologico e scaturisce dalla contrapposizione ancora irrisolta, e forse irrisolvibile, tra due modi diversi di concepire la causalità, una di tipo *sequenziale* e una di tipo *generativo***. Per alcuni Perotti e Teoldi si sono resi colpevoli di un'offesa al pluralismo metodologico, per aver spezzato troppo frettolosamente una lancia a favore dell'uso di un solo "metodo". Avrebbero potuto facilmente evitare questo tipo di critiche invocando l'uso di metodi misti, o prestando una qualche attenzione all'esistenza della *theory-based evaluation*, che ha un potenziale cognitivo secondo i suoi sostenitori non inferiore a quello dell'analisi controfattuale. Anzi, i metodi *theory-based* sono giustificati soprattutto laddove il metodo controfattuale entra in crisi a causa della **complessità** del programma e dei **meccanismi causali** sottostanti: ci si aspetta che l'applicazione di questi metodi alternativi, a parità di esito cognitivo, sia ancora più sfidante intellettualmente dell'analisi **controfattuale stessa**. "*Intellectually enormously challenging*", l'hanno definita due tra i suoi più quotati esponenti, in quanto richiede "*advanced theoretical understanding, abilities to design research and techniques for data analysis*."***

Gli autori dell'articolo su *lavoce.info* però si sono anche resi colpevoli di un secondo misfatto, molto più grave agli occhi di un numero più grande di critici. Nel denunciare che l'imperatore è senza vestiti, ovvero che le valutazioni dei fondi strutturali condotte in questi anni sono servite davvero a poco in termini decisionali, Perotti e Teoldi hanno anche sottoposto ad una severa reprimenda la comunità di



valutatori cresciuta all'ombra di tali fondi (il "sottobosco", come viene chiamato nell'articolo).

Mi pare evidente che il pluralismo metodologico c'entri poco nell'autodifesa dell'AIV, e che i tantissimi numeri inutili sui fondi strutturali non siano certo il prodotto di alcuna *theory-based evaluation*, semmai molto più il prodotto degli incentivi perversi che hanno in questi anni favorito la crescita del "sottobosco".

In verità, una qualche insofferenza verso la vacuità dei tanti indicatori generati e delle migliaia di pagine scritte nei rapporti di valutazione, occasionalmente affiorava, però non è mai arrivata ad una denuncia esplicita, ad una ferma presa di posizione. Fino a che qualcuno, dall'esterno della comunità, non ha mostrato pubblicamente che l'imperatore è senza vestiti. Ma come si è arrivati a questa situazione? Come è stato possibile fingere per così tanto tempo che quel tipo di valutazione servisse davvero a qualcosa? Occorre capire a quali incentivi (perversi) rispondono i valutatori indipendenti dei programmi regionali (POR) e nazionali (PON) dei fondi strutturali.

Gli incentivi perversi del valutatore indipendente

Nell'ambito dei fondi strutturali, ogni "valutazione" viene messa a bando dalla competente Autorità di Gestione, seguendo in buona parte le regole imposte dalla Commissione. I capitolati di questi bandi sono confezionati tagliando/incollando brani interi dai già fumosi documenti comunitari, che devono andare bene a tutti i 28 stati membri, dall'Estonia al Portogallo, dalla Svezia a Malta. I capitolati finiscono per essere pieni zeppi di retorica e *aria fritta* e di domande tanto ambiziose da sfiorare la megalomania, mascherata dietro altisonanti indicatori; sono però privi di quesiti a cui si possa dare una risposta rilevante per le decisioni con i dati a cui il valutatore indipendente ha un reale accesso.

Chi partecipa a queste gare sa bene che la sua offerta sarà giudicata da qualcuno a sua volta poco interessato ai contenuti e ai risultati della "valutazione" stessa. Chi partecipa tende quindi a investire poche risorse sulla qualità della proposta tecnica e soprattutto per proporre analisi con un impianto rigoroso. Piuttosto massimizza il numero di bandi a cui partecipa, sfruttando economie di scala e sperando di spuntarla con i ribassi sulle offerte economiche. Questi incentivi

perversi condannano le "valutazioni" a produrre quantità frastornanti di numeri, per lo più ricavate dai dati di monitoraggio, che non dicono nulla, se non, nel migliore dei casi, descrivere ciò che è stato fatto. Certo non a cogliere gli effetti dei soldi spesi, nonostante l'abuso di termini quali "effetti" o "impatti".

Questa comunità di produttori di numeri all'ingrosso è, più o meno consapevolmente, la *foglia di fico* di un sistema politico-amministrativo messo in piedi per "spendere tutto" senza mai mettere in discussione alcuno degli interventi pagati dai fondi strutturali, cioè senza mai "*sapere che effetti producono.*"

Perotti e Teoldi hanno messo brutalmente sul tappeto una domanda che i *policy-maker* si pongono raramente, e cioè "*come facciamo a sapere se i benefici di questi progetti superano i costi per la collettività?* [...] *Come in tutte le questioni di economia non esiste e non esisterà mai una risposta certa alla domanda di partenza. Ci sono però modi più o meno sofisticati per cercare di avvicinarsi ad una risposta ragionevole.*"

Una risposta credibile a una domanda del tipo "sono soldi ben spesi?": questo ci si sarebbe

aspettato dalle centinaia di rapporti di valutazione prodotti dai valutatori indipendenti. Questo non è successo. Perché questi rapporti non raggiungono conclusioni rigorose sugli effetti dei fondi strutturali? Sarebbe stato sufficiente forzare il metodo controfattuale (o addirittura i più sfidanti e ambiziosi metodi *theory-based*) a tutti i POR e PON d'Italia?

La risposta è ovviamente negativa per un complesso di motivi, che però possono essere ricondotti a due: (i) cosa sanno veramente (fare) i valutatori italiani che vivono nel sottobosco di perottiana definizione, e i loro committenti nelle autorità di gestione, e (ii) cosa veramente motiva i *policy-maker* italiani a utilizzare evidenza sugli effetti delle politiche che promuovono?

Senza voler pregiudicare il futuro, che potrebbe rivelarci sviluppi oggi inaspettati – con tempi che lasciamo all'ottimismo o pessimismo del singolo lettore – possiamo solo guardare alla situazione presente e avanzare forse qualche ipotesi a come ci si è arrivati. L'applicazione su ampia scala di metodi rigorosi, siano essi di natura controfattuale o *theory-based*, è oggi impossibile in Italia, semplicemente perché le persone formate a pensare in modo rigoroso alla valutazione delle politiche sono pochissime nel "sottobosco", e praticamente **inesistenti** nelle autorità di gestione, PON o POR che siano.

In altri paesi queste figure vengono formate all'interno di dottorati con un'enfasi sui metodi quantitativi avanzati. Una delle ragioni del rapido diffondersi dell'approccio controfattuale nei governi dei paesi anglosassoni (e dell'America Latina, in questo fortemente influenzati dall'esempio statunitense) e nelle organizzazioni internazionali

come la Banca Mondiale e l'OCSE è la storicamente forte presenza in queste organizzazioni di *Ph.D. trained economists*, che possono più facilmente essere convinti (o illusi) della scientificità del metodo controfattuale di quanto possa esserlo un dirigente laureato in Giurisprudenza di una regione italiana. Una massiccia iniezione di *Ph.D. trained economists* nella pubblica amministrazione italiana pare oggi meno probabile del proverbiale cammello che passa per la cruna di un ago. Così come appare altrettanto improbabile oggi un improvviso risveglio di interesse da parte dei *policy-maker* per l'evidenza sugli effetti delle politiche pubbliche, siano esse co-finanziate o meno dai fondi strutturali (il perché lo lasciamo agli studenti come esercizio da svolgere a casa).

Quello contenuto nei due paragrafi precedenti potrebbe suonare come un pessimistico invito a rassegnarsi ad uno status quo insoddisfacente. Ma l'insoddisfazione non deve necessariamente tradursi in rassegnazione o inazione: sono anzi convinto che alcune cose utili possano essere fatte nell'attesa di tempi migliori. Una prima cosa da fare è smettere con la contrapposizione tra controfattuale e *theory-based*, che se può avere qualche fascino quando è gestita dagli *original thinkers*, diventa un po' stucchevole quando è affidata a menti molto meno sapienti come quella di chi scrive. Personalmente, trovo convincenti le argomentazioni di Stern *et al.* (2012) quando scrivono che: *Counterfactual [method]s answer contingent, setting-specific causal questions "did it work there and then" and cannot be used for generalization to other settings and timeframes, unless they are accompanied by more fine-grained*

knowledge on the causal mechanisms actually operating within the process leading from potential cause to effect." Ne sarei ancora più convinto se si potessero vedere molti esempi di tale "*fine-grained knowledge on the causal mechanisms*" che consentano alle stime controfattuali di essere generalizzate.

Una seconda cosa da fare nell'attesa di tempi migliori è non accontentarsi dei risultati raggiunti, ma mettere autocriticamente in discussione tali risultati. Sono convinto che il numero monografico della RIV, il n. 58 di imminente uscita – dedicato alla valutazione d'impatto delle politiche in Italia e focalizzato sull'analisi controfattuale – rappresenti una grande ragione di soddisfazione per i suoi autori e curatori, ma anche un'ottima occasione per una riflessione critica per tutta la comunità dei valutatori.

Lo spunto per questa riflessione mi viene dal discorso di accettazione del premio intitolato a Peter Rossi****, in cui l'economista americano Larry Orr ammette molto candidamente: "*The great methodological battle of the last 40 years – how to ensure internal validity – is essentially resolved*"****, mentre su altri due temi cruciali per l'utilizzabilità delle stime di impatto, e cioè "*ensuring external validity*" come pure "*dealing with the very low success rate of social interventions when tested rigorously*", la partita è ancora tutta da giocare.

A me pare fuori di dubbio che, se l'obiettivo della valutazione è quantificare effetti attribuibili ad una politica, l'approccio controfattuale rappresenti il necessario punto di partenza, se non altro perché rappresenta un enorme patrimonio di esperienze,



consapevolezze ed elaborazioni rigorose accumulate in 40 anni da una ampia comunità di studiosi. Va tuttavia anche riconosciuto come questo enorme patrimonio di strumenti *non fornisca risposte a tutte le domande rilevanti per le decisioni pubbliche*.

Su questo mi trovo d'accordo con Nicoletta Stame quando scrive *"only die-hard followers of the counterfactual can claim to be mainly interested in internal validity, without realizing that this runs against their desire to offer the policy-maker a strong argument for taking far-reaching decisions"* (Stame, 2012, p. 380).

Pur con argomentazioni molto diverse, sia Nicoletta Stame sia Larry Orr sostengono che al di là della validità interna, la partita sia ancora in gran parte da giocare. Si potrebbe anche argomentare che dovrebbero entrare in campo a pieno titolo politologi, psicologi, esperti di comunicazione e sociologi. Una volta stabilita la validità interna delle stime degli effetti, resta infatti ancora molto da fare, e chi si occupa per mestiere di costruire evidenze robuste in genere non sa necessariamente come muoversi per spingere i decisori pubblici

verso l'utilizzo di quelle stesse evidenze. Semplicemente, ci vogliono altre competenze e sensibilità, ben al di là della *proficiency* econometrica.

Va riconosciuto che una caratteristica comune degli studi del monografico della RIV è l'attenzione quasi esclusiva alla **validità interna** delle stime prodotte, cioè al fatto che le stime degli effetti ottenute con *quel* particolare campione analizzato non siano distorte dalla presenza di *selection bias*. Gli autori dedicano scarsa attenzione all'utilizzo futuro dei loro risultati, vuoi perché abituati da sempre a trovare scarsa attenzione da parte del decisore pubblico o perché trovano motivazioni e gratificazioni professionali altrove (la carriera accademica dipende dalle pubblicazioni scientifiche prodotte, non certo da quanto si è riusciti a spiegare al dirigente o all'assessore di turno il senso e l'utilità delle analisi condotte).

Tra gli studi inclusi nel numero monografico della RIV ce n'è uno solo che fa un uso (almeno parziale) del disegno randomizzato, che secondo molti dovrebbe rappresentare il *gold standard* della valutazione. Conti

et al. (2015) studiano gli effetti di un intervento di prevenzione del *drop-out* negli istituti professionali (*reducing dropout in the early grades of vocational high schools*.) Dopo un'analisi che mostra sia la superba competenza statistica degli autori e l'esiguità della base empirica su cui l'analisi si basa (53 classi di cui 18 trattate e 35 di controllo), ma mostra al contempo quanto possa diventare alta la barriera alla comprensione di studi di una tale complessità, la conclusione che gli autori traggono dai loro sforzi erculei è abbastanza misera: *"The results suggest that Innovare has positive effects. Although the estimated effects are small, there is some evidence that [the programme ...]. Innovare decreases the probability of failing and dropping out, reduces the absence rate...."* (Conti *et al.* 2015, p. 22) *The effects of the Innovare program are promising, although they are small in size and not highly significant. Also our results suggest that implementing similar projects to analogous contexts may be worthwhile. The importance to invest in this type of programs is further stressed by the results from the qualitative*

evaluation. The focus groups pointed out that the program contributed to improve the behavior of students in the classroom and their attitude towards learning.”

Una situazione che può simultaneamente produrre un'impeccabile validità interna senza alcuna credibile validità esterna si verifica quando le stime dell'impatto sono ottenute reclutando *volontari* interessati a partecipare ad una determinata esperienza: l'obiettivo della valutazione è prevedere l'effetto dell'intervento una volta che sia reso accessibile a tutti o addirittura sia dichiarato *obbligatorio*. Un esempio sono le innovazioni didattiche, la cui efficacia è testata reclutando singole scuole o singoli insegnanti che si offrono volontari per implementare quella sperimentazione. La logica controfattuale richiede che ad un sotto-gruppo *scelto a caso* di volontari sia successivamente negata l'opportunità di partecipare alla sperimentazione. Anche ammesso che tutti si comportino secondo le consegne ricevute (*perfect compliance*) l'effetto dell'innovazione sui volontari potrebbe essere molto diverso dagli effetti prodotti quando essa viene generalizzata. La tipica preoccupazione è che l'esperimento suggerisca effetti positivi che poi non si materializzano quando la sperimentazione è estesa a tutte le scuole. La credibilità di questa generalizzazione è molto scarsa. Questa è un'area senza dubbio ancora poco esplorata, anzi spesso ignorata. Un buon proposito è tentare di rispondere sempre alle classiche domande sulla validità esterna: a quali condizioni sono le stime di impatto generalizzabili a diversi contesti, a gruppi con livelli diversi di coinvolgimento, o semplicemente ad una scala di

livello superiore. Questa attenzione renderebbe quanto ottenuto più utilizzabile per le decisioni di policy – *sempre se vengono davvero tempi migliori*.

Il volume monografico della RIV offre anche esempi in cui appare minore la distanza tra le stime degli effetti e un possibile utilizzo nel disegno delle politiche. Ad esempio, Augusto Cerqua e Guido Pellegrini si chiedono: *“Conviene incentivare le piccole imprese piuttosto che le grandi?”* e provano a dare una risposta utilizzando la versione più avanzata del *regression discontinuity design*. Il seguente è un estratto delle conclusioni: *“L'analisi precedente segnala la presenza di un'evidente eterogeneità nell'intensità e significatività degli effetti delle agevolazioni erogate dalla L. 488 per dimensione d'impresa. Il segno di tali effetti è coerente con quanto previsto dalla letteratura teorica: le imprese più piccole subiscono in misura maggiore i problemi di razionamento dal lato del credito e quindi reagiscono agli incentivi attivando progetti con una maggiore addizionalità rispetto alle imprese di maggiore dimensione. [...] Nel complesso, gli effetti della L. 488 sono in linea con almeno uno degli obiettivi previsti dai policy-maker: le imprese agevolate investono più di quelle non agevolate, creano più capitale tangibile e vendono in misura maggiore delle non agevolate. Questi effetti sono concentrati tra le micro e piccole imprese, sostenendo l'ipotesi che gli incentivi effettivamente riescano a sopperire a evidenti imperfezioni nel mercato del credito.”*

Secondo Larry Orr, l'altra minaccia concreta all'utilizzabilità delle stime controfattuali è poi l'ormai *lunga litania di effetti statisticamente significativi ma*

dalle irrisorie implicazioni di policy. Quello che Orr definisce *the very low success rate of social interventions when tested rigorously*, sia chiaro, non è una conseguenza del metodo controfattuale, e potrebbe non riprodursi con la stessa frequenza in contesti diversi da quello statunitense. Ma in presenza di questo fenomeno, è facile che vengano considerati poco utili i risultati conoscitivi modesti prodotti da valutazioni a volte anche molto costose.

Quali sono le soluzioni al problema della frequente irrilevanza pratica delle stime controfattuali? Innanzitutto, produrle in modo sistematico e non episodico e quasi aneddótico, come si fa ora nel nostro paese, per cui neppure sappiamo se un problema analogo esiste anche da noi. La seconda, poco esplorata, è quella di non fermarsi alla significatività statistica ma trarre le spesso sorprendenti implicazioni che hanno semplici calcoli di costo-efficacia, cioè di *costo per unità di impatto*.

Un esempio di tali calcoli di costo-efficacia è il seguente. Poniamo che un RCT realizzato secondo lo stato dell'arte indichi che un certo tipo di corsi di formazione professionale per disoccupati abbia prodotto il seguente risultato. Dopo N anni dall'inizio dei corsi di durata annuale, il tasso di occupazione dei controlli, esclusi mediante sorteggio dalla frequenza dei corsi sia il 45%, mentre tra coloro che hanno frequentato i corsi sia del 50%. L'impatto (o effetto) della formazione in quel caso è di 5 punti percentuali, ovvero l'11% rispetto alla situazione che si sarebbe determinata in assenza di intervento. Poniamo che l'esperimento abbia coinvolto circa 10.000 studenti: con un campione di tali dimensioni le



stime saranno abbondantemente significative. *Statisticamente però*. Ma cosa se ne può fare di una stima di un effetto di 5 punti percentuali l'Assessore al Lavoro e alla Formazione professionale di una qualunque regione, o il suo Direttore Generale? Quali decisioni possono essere influenzate da tale evidenza? Espandere la platea dei beneficiari, sperando di sfruttare possibili economie di scala, oppure restringerne il numero dei formati tentando di mirare meglio l'intervento, o addirittura abolirlo perché troppo costoso?

Quando un effetto è positivo, la parola "abolire" sembra fuori luogo. Facciamo due conti. Poniamo che il costo del corso sia di 2.000 euro per studente e che sia stato frequentato da 10.000 studenti. Quindi il costo per la Regione e per l'FSE sarà di 20 milioni di euro. Qualche consigliere di opposizione potrebbe far notare che a beneficiare in modo netto dalla formazione saranno solo quei 500 studenti che senza il corso non avrebbero trovato lavoro (cioè il 5% di 10.000). Venti milioni diviso 500 fa 40.000 (quarantamila euro) per studente occupato grazie alla formazione, cioè 20 volte il costo per studente

iscritto. Vista così, la formazione non sembra più un affare per la collettività. È difficile che *i benefici di questi progetti superano i costi per la collettività*, come chiedono che si verifichi Perotti e Teoldi, anche quando i benefici futuri siano attualizzati usando un appropriato tasso di sconto. Diverse ovviamente sarebbero le conclusioni se l'impatto fosse di 30 punti percentuali, differenza ottenuta sottraendo dal 50% di occupati tra i formati un molto più misero 20% di occupati tra i controlli. Questo abbasserebbe il costo di un occupato addizionale a 20 milioni diviso 3.000 cioè un molto più accettabile 6.667 euro, poco più di 3 volte il costo per studente frequentante. In questo caso sarebbe più facile argomentare a favore di un'espansione dell'intervento. Tuttavia, questo scenario è assai raro; molto più frequente è lo scenario dell'impatto significativo ma ostinatamente piccolo, da cui è difficile trarre implicazioni di *policy* tali da indurre il *policy-maker* a cambiare le due decisioni. Le precedenti considerazioni mettono in evidenza come il metodo controfattuale, *gold* o *bronze standard* che sia, non risolve tutti i problemi conoscitivi che si pongono nei confronti

dell'efficacia di una politica pubblica, né mette tutti d'accordo, soprattutto quando un'evidenza debole lasci più ampi spazi all'interpretazione. L'applicazione della logica controfattuale delegittima tuttavia le affermazioni più avventate, quali quelle che si potrebbero fare se non ci fosse un gruppo di controllo. Ad esempio, che grazie alla formazione ricevuta più della metà dei formati ha trovato lavoro. Nella consueta metafora, il controfattuale non consente di stabilire *in modo oggettivo* se il bicchiere è mezzo vuoto o mezzo pieno. Ma almeno garantisce che si stia tutti parlando dello **stesso bicchiere**.

I termini su cui provare a costruire almeno una **coesistenza pacifica**, se non una **collaborazione nella diversità**, tra controfattualisti e *theory-based* sarebbero i seguenti: (i) il *primo passo* (che soffre di limiti di cui sono tutti consapevoli e che non è necessario qui ribadire) verso la valutazione di una politica è la quantificazione dei suoi effetti causali, che (ii) nella gran parte dei casi hanno **validità interna**, ma chiedono (iii) essere compresi nei loro **meccanismi** e (iv) resi in una forma che ne consenta l'**utilizzo** per prendere decisioni.

Quando si valuta una politica si parte *per default con una prima analisi controfattuale, consapevole fin dall'inizio* del fatto che da sola è insufficiente. Nel quadro di tale coesistenza, ciascuno può mantenere le sue convinzioni e le sue riserve di tipo epistemologico, ma dal punto di vista operativo per valutare una politica e i suoi effetti il primo passo sarebbe tendenzialmente di natura controfattuale. Almeno in linea di principio. Tutto ciò nell'attesa, *ça va sans dire*, di tempi migliori.

Scrivendo Tom Cook in un saggio apparso sul *Journal of Multidisciplinary Evaluation*, (2006, p. 6) la retorica del 'Gold Standard' va messa nella giusta prospettiva: “[...] *studies with different methods so long as different types of questions are asked. One truism of theory of method is that different questions usually require different methods if these questions are to be answered well. Even a good experiment requires attention to theory, sampling, qualitative data collection, value analysis and the like. So even a preference for experiments requires a commitment to multi-method research.*”

Riferimenti bibliografici

Cerqua A. and Pellegrini G., (2015) “Conviene incentivare le piccole imprese piuttosto che le grandi? Un’analisi basata sul multiple regression discontinuity design” *Rassegna Italiana di Valutazione*, n. 58, di prossima pubblicazione.

Conti E., Duranti S., Mattei A., Mealli F., and Sciclone N., (2015), “The effects of a dropout prevention program on secondary students’ outcomes”, *Rassegna Italiana di Valutazione*, n. 58, di prossima pubblicazione.

Cook, T.D. (2006) “Describing what is Special about the Role of Experiments in Contemporary Educational Research: Putting the ‘Gold Standard’ Rhetoric into Perspective”, *Journal of Multidisciplinary Evaluation*, 3 (6), pp. 1-7.

Cook *et al.* (2010) “Contemporary Thinking About Causation in Evaluation: A Dialogue With Tom Cook and Michael Scriven”, *American Journal of Evaluation*, 31(1), pp. 105-117.

Orr L. (2014) “Beyond Internal Validity”, *Peter Rossi Award Lecture delivered at the Association For Public Policy Analysis and Management, 2014 Annual Conference, November 2014*.

Pawson, R. and N. Tilley (1997) *Realistic Evaluation*, Sage Publications, Thousand Oaks.

Pawson R. and N. Tilley (2004), *Realistic Evaluation*, in Matthieson S. (ed.), *Encyclopaedia of Evaluation*, Beverly Hills, CA, Sage, 2004.

Perotti R. and Teoldi F. (2014) “Il disastro dei fondi strutturali”, *Lavoce.info*, 3 luglio 2014.

Stame, N. (2012) “What Doesn’t Work? Three Failures, Many Answers”, *Evaluation*, 16(4), pp-371–387.

Stern, E., Stame N., Mayne, J. , Forss K. , Davies, R. and Befani B. (2012) “Broadening the Range of Designs and Methods for Impact Evaluations”, *Final report of a study commissioned by the UK Department for International Development, April 2012*.

Note

*Intendo ringraziare Marco Sisti, Ugo Trivellato e Barbara Romano per gli utili commenti e suggerimenti su versioni precedenti di questo lavoro.

** Ci sono molti altri modi di caratterizzare questa contrapposizione: per una stimolante rivisitazione di questi temi si veda il vivace scambio tra Tom Cook e Michael Scriven, riportato in Cook *et al.* (2010).

***Pawson and Tilley (2004, p. 20).

****Il premio Peter Rossi viene conferito ogni due anni *for outstanding contribution to evaluation methodology*. In passato hanno avuto il premio Tom Cook, Judith Gueron, Rebecca Maynard, Howard Bloom, Frank Mosteller e Rob Hollister.

*****Notate come Orr dica *resolved* e non *won* riferendosi alla battaglia metodologica, in quanto il “nemico” non è una particolare scuola di pensiero, bensì le difficoltà concettuali poste da una realtà sociale complessa.

XVIII Congresso nazionale

17-18 aprile 2015

DIPARTIMENTO DI SCIENZE DELLA FORMAZIONE

Università di Genova

Corso Andrea Podestà, 2 Genova

CALL FOR PAPERS

VALUTAZIONE E EQUITÀ SOCIALE

La valutazione come fattore di cambiamento delle politiche pubbliche e strumento per il contrasto delle disuguaglianze sociali

Nell'anno internazionale della valutazione, il tema dell'equità sociale è particolarmente sentito. La crisi economica globale e l'accrescersi delle disuguaglianze sociali nel mondo richiamano l'attenzione sulle politiche che mirano a garantire a tutti l'accesso al godimento dei diritti di base (alla vita, salute, istruzione) e a contrastare le disparità persistenti tra i gruppi sociali, i territori e i paesi nel mondo. Tuttavia, le politiche possono dar luogo a effetti avversi e generare nuove disuguaglianze. Talvolta, in presenza di leggi, regolamenti e prescrizioni contrari a qualsiasi forma di discriminazione sociale, etnica, di genere e di classe, si possono verificare e perpetuare iniquità che minano la libertà delle persone, le possibilità di sviluppo dei territori e il funzionamento efficiente e democratico delle istituzioni. Anche all'interno della "macchina" amministrativa e dei processi organizzativi possono nascere disuguaglianze. In tali circostanze, la valutazione è chiamata ad intercettare gli esiti indesiderati e a contribuire al ripristino di condizioni di equità. La stessa valutazione non è immune da potenziali effetti iniqui: talvolta la selezione dei metodi (quantitativi e qualitativi) e degli approcci (controfattuale, theory-based, costi-benefici, partecipativi) può essere influenzata da dinamiche di potere e da conflitti valoriali che caratterizzano la produzione di conoscenza valutativa e le conclusioni e i giudizi che ne derivano.

Nel 2015 - anno internazionale dedicato alla valutazione - obiettivo del XVIII Congresso nazionale dell'Associazione Italiana di Valutazione è esplorare le dimensioni dell'equità sociale nella valutazione delle politiche pubbliche. Il Congresso intende essere l'occasione per esaminare gli approcci, i metodi e gli indicatori valutativi utili a comprendere in che modo le diverse forme di disuguaglianza, che riguardano sia le condizioni di partenza sia gli esiti delle politiche attuate, possano essere superate. La sfida delle sessioni tematiche in cui si declinerà il tema del congresso è quella di affrontare i diversi significati assegnati al concetto di equità sociale dal disegno e dalla pratica di valutazione. Il tema dell'equità sociale è, infatti, trasversale ai molti settori nei quali i valutatori operano, come:

- 1 - Le politiche formative (istruzione, formazione professionale)
- 2 - L'università e le politiche per la ricerca scientifica e l'avanzamento tecnologico
- 3 - La pubblica amministrazione
- 4 - La sanità
- 5 - Gli enti locali/lo sviluppo locale
- 6 - La cooperazione e gli aiuti umanitari
- 7 - Le politiche sociali e di inclusione sociale
- 8 - Le politiche attive e passive del lavoro
- 9 - Le politiche urbanistiche e di rigenerazione urbana
- 10 - La gestione di programmi complessi
- 11 - Le politiche di genere
- 12 - I processi di innovazione sociale
- 13 - Le politiche europee

Il Comitato scientifico-organizzativo del XVIII Congresso nazionale dell'Associazione Italiana di Valutazione invita a inviare proposte di sessioni e relazioni da presentare nell'ambito del congresso, che si terrà a Genova, il 17 e 18 aprile 2015 presso l'Università di Genova, Dipartimento di Scienze della Formazione.

Come nei precedenti congressi, il Comitato Scientifico-Organizzatore propone un'ampia selezione di temi da approfondire in sessioni dedicate, tavole rotonde, gruppi di lavoro che coinvolgano esperti e portatori di interessi che condivideranno diversi saperi e esperienze. Oltre a quelli sopra indicati, si possono proporre altri temi, oppure suggerire di associare uno o più temi tra loro. I contributi potranno proporre, sia dal punto di vista teorico che da quello pratico, studi, riflessioni, argomentazioni utili a mettere in evidenza il ruolo della valutazione rispetto al tema dell'equità e della lotta alle disuguaglianze.

Gli abstract sono soggetti a valutazione anonima da parte del Comitato scientifico nominato dal direttivo dell'AIV che comunicherà l'accettazione entro il 16 marzo 2015.

I criteri di valutazione dei contributi sono i seguenti: 1. Rilevanza dei contenuti rispetto al focus centrale del Congresso. 2. Chiarezza dell'obiettivo e dei sub-obiettivi. 3. Chiarezza della metodologia. 4. Originalità rispetto ai paradigmi e alle metodologie disciplinari consolidate e innovatività delle tesi discusse e/o delle soluzioni operative proposte sulla base di adeguata argomentazione. I contributi eccellenti potranno essere menzionati nel corso del Congresso e, successivamente, proposti per la candidatura alla pubblicazione sulla RIV, opportunamente rivisti e adattati. Coloro che ne faranno richiesta potranno pubblicare i contributi accettati al Congresso saranno sul sito valutazioneitaliana.it

CALENDARIO DELLE SCADENZE

- Apertura call for paper	9 dicembre 2014
- Deadline per la presentazione degli abstract	13 febbraio 2015
- Comunicazione degli abstract accettati	16 marzo 2015
- Invio dei paper ai coordinatori di sessione	31 marzo 2015

COMITATO SCIENTIFICO/ORGANIZZATORE

Maria Carmela Agodi (Università di Napoli "Federico II")

Daniele Bondonio (Università del Piemonte Orientale "Amedeo Avogadro")

Daniela Congiu (Mixura)

Michela Freddano (Invalsi)

Veronica Lopresti (Sapienza Università di Roma)

Mita Marra (Università di Salerno)

Francesco Mazzeo (Università di Catania)

Erica Melloni (IRS, Milano)

Mauro Palumbo (Università di Genova)

Carlo Pennisi (Università di Catania)

Gabriele Tomei (Università di Pisa)

Claudio Torrigiani (Università di Genova)

Alberto Vergani (Istituto Italiano di Valutazione)

SEGRETERIA ORGANIZZATIVA

Segreteria AIV

Tel. 081.2452273 - Fax 081.7648924

segreteria@valutazioneitaliana.it

Ricordiamo inoltre ai soci AIV le seguenti iniziative che si svolgeranno nei prossimi mesi:

Call for paper: Conferenza internazionale sulla politica pubblica 2015

Nel corso degli ultimi 30 anni, la ricerca sulla politica pubblica è diventato oggetto di studio interdisciplinare, contribuendo a comprendere i fenomeni legati al public policy-making.

Al fine di continuare a riunire tutti i ricercatori che si occupano di politiche pubbliche e i loro diversi approcci e tematiche, vi invitiamo a partecipare alla seconda Conferenza Internazionale sulla Politica Pubblica che si terrà presso l'Università Cattolica del Sacro Cuore, a Milano, Italia dal 1 al 3 luglio 2015.

Durante questa conferenza, si offrirà l'opportunità sia ai ricercatori junior che a quelli senior di presentare e discutere le nuove ricerche, intuizioni teoriche, concettuali e metodologiche e risultati empirici attraverso un simposio di panel e laboratori selezionati.

La conferenza coinvolgerà anche relatori invitati e discussioni plenarie.

Ottava conferenza annuale ESPAnet Italia 2015**Welfare in Italia e welfare globale: esperienze e modelli di sviluppo a confronto****Fisciano (SA), 17-19 settembre 2015****Tema della conferenza**

Con la crisi economica globale, le politiche di austerità dell'Unione Europea e le persistenti carenze amministrative, i nuovi rischi sociali si sono acuiti mettendo sotto pressione il welfare italiano, in particolare nel Mezzogiorno. L'attenzione delle politiche sociali si è concentrata sui disagi più urgenti emersi negli ultimi anni nella cornice politico-istituzionale europea, già sottoposta a un profondo processo di trasformazione. Sono rimaste sullo sfondo sia la valutazione dei sistemi di protezione sociale emergenti nel resto del mondo, sia la ricostruzione storica dei periodi di povertà e di prosperità che si sono alternati nel nostro paese nella lunga durata — prospettive analitiche che sarebbero utili a formulare politiche più mirate e innovative.

Obiettivo dell'ottava conferenza di ESPAnet Italia 2015 è integrare l'analisi diacronica delle politiche sociali con una più ampia considerazione dei sistemi di welfare emergenti in altri contesti nazionali europei ed extra-europei. L'idea che sottende la conferenza consiste nell'esaminare l'esperienza italiana nel tempo, rispecchiandola nei sistemi di welfare in corso di sviluppo, in particolare nei paesi con una più rapida crescita economica. Studiosi internazionali, italiani ed europei, operatori del sociale, studenti e ricercatori discuteranno sul passato e sul futuro del nostro welfare: presenteranno e confronteranno i casi più interessanti del Nord e del Sud del mondo e del nostro paese.

Per maggiori informazioni visita il sito:

<http://www.espanet-italia.net/conferenza2015.html>



Per associarsi ad AIV, le quote d'iscrizione sono:

-100,00 euro, soci ordinari

-50,00 euro, soci giovani, a norma di Statuto e Regolamento sotto i 30 anni di età e i dottorandi (senza limiti di età), iscritti ai Master patrocinati AIV.

-150,00 euro, soci amici dell'AIV

Effettuare il pagamento della quota associativa tramite bonifico bancario versando la quota associativa sul c/c di Banca Prossima IBAN IT 41 L033 5901 6001 0000 0062 397

Importante!!! Nella causale del bonifico specificare il proprio nome e cognome