

Alberto Martini\*  
Marco Sisti\*\*

**Indicatori o analisi di performance?  
Implicazioni dell'esperienza statunitense  
di *performance measurement***

\*Docente di Statistica Economica, Università del Piemonte Orientale e Direttore, Progetto Valutazione  
\*\*Ricercatore, Progetto Valutazione

## Introduzione

L'uso di indicatori per misurare e valutare la performance di organizzazioni pubbliche si sta diffondendo rapidamente in Italia. L'ultima forte spinta in questo senso viene dal decreto legislativo 286 del 1999, che all'art. 4 impone a tutte le amministrazioni di dotarsi, quali strumenti per il controllo di gestione, di indicatori di efficienza, efficacia ed economicità. Anche in ambiti che non si richiamano al controllo di gestione, quali i programmi di intervento comunitari, è molto comune il richiamo all'uso di indicatori come strumento di valutazione. Il Quadro Comunitario di Sostegno del Fondo Sociale Europeo per il periodo 2000-2006 contiene nutrite batterie di indicatori con cui i titolari dei programmi operativi sono chiamati a rendere conto dell'uso dei fondi comunitari. Pare esistere un consenso pressoché unanime sul fatto che la disponibilità di indicatori sia fondamentale per osservare e quindi migliorare la performance della pubblica amministrazione.

La popolarità degli indicatori di performance non è un fatto solo italiano. Molti dei tentativi condotti in altri paesi per razionalizzare la gestione della pubblica amministrazione ruotano attorno alla costruzione di sistemi di misurazione e valutazione. Dopo le prime pionieristiche esperienze negli anni '60, come il *Planning Programming Budgeting System* statunitense, questi tentativi hanno avuto un notevole sviluppo durante gli anni '70 e '80, soprattutto nei paesi anglosassoni. Il punto di partenza di tali sistemi tende sempre ad essere la necessità di supplire, all'interno del settore pubblico, alla mancanza degli strumenti fondamentali di misurazione (del prodotto) e valutazione (dei risultati) di cui dispongono le aziende private: i prezzi di vendita ed i profitti.

Agli occhi di coloro che si cimentano nel tentativo di migliorare la pubblica amministrazione, qualsiasi strumento analitico che prometta di "misurare e valutare" ciò che la pubblica amministrazione produce tende quindi ad avere un grande *appeal*. Ed è anche naturale che gli strumenti più semplici e rassicuranti incontrino il maggior favore. Gli indicatori di performance soddisfano entrambi questi requisiti: sono facilmente comprensibili, trattandosi di semplici quantificazioni di fenomeni o di rapporti tra due quantità; paiono offrire la possibilità di "valutare" direttamente qualsiasi dimensione di successo dell'attività pubblica, sia essa definita come efficienza, efficacia, economicità, risultato, esito, impatto, rendimento, qualità. Con un limitato impegno analitico e un ridotto sforzo di misurazione essi sembrano produrre gran parte dei risultati conoscitivi necessari per sostenere l'opera di razionalizzazione della pubblica amministrazione.

Presentiamo qui alcune riflessioni critiche sull'uso e l'utilizzabilità di questo strumento a fini di valutazione delle organizzazioni pubbliche. Il nostro scopo non è deplorare l'uso degli indicatori, bensì stimolare una maggiore consapevolezza dei limiti conoscitivi di questo semplice strumento analitico e delle cautele che dovrebbero circondare il suo uso. Nel corso del lavoro useremo il concetto di "analisi di performance" per sottolineare l'idea che la costruzione degli indicatori non debba essere fine a sé stessa, bensì debba essere vista come uno degli elementi di un disegno analitico più complessivo. Elementi di questo disegno più complessivo emergono chiaramente dall'esame delle esperienze statunitensi illustrate più avanti.

Non siamo gli unici ad avanzare cautele su un uso troppo "facile" degli indicatori. Un valutatore canadese, Burt Perrin (1998), sostiene che «*performance measures are frequently irrelevant—or worse. Used in isolation, they invariably reward and encourage the wrong activities and result in less, rather than more, attention to outcome and quality*». Simile scetticismo è stato espresso da tre studiosi francesi (Dejean, Bourassin e Mortreux, 1998), i quali notano con disappunto come «*dans certains cas, c'est comme si la question de l'évaluation se réduisait à celle des indicateurs*». Più avanti affermano invece che «*la formulation d'indicateurs n'est qu'une étape parmi d'autres de la construction d'un dispositif d'évaluation*». In entrambi gli articoli gli autori esprimono riserve su un uso poco meditato degli indicatori e danno dei consigli pratici per migliorarne l'utilizzo. Questo è anche lo spirito che anima il nostro lavoro.

Gli indicatori vengono utilizzati in una gran varietà di situazioni che a diverso titolo si richiamano alla valutazione. In questo articolo ci concentriamo sugli indicatori utilizzati come strumento per valutare la performance di singole organizzazioni pubbliche o sistemi di organizzazioni. Con "performance" intendiamo l'insieme delle caratteristiche desiderabili dell'operare di un'organizzazione (i suoi costi; la qualità delle prestazioni; i volumi di attività; la ricaduta che ottiene sull'ambiente esterno verso cui agisce). Per analisi di performance intendiamo lo sforzo di dare dei giudizi sui vari aspetti dell'operare di

un'organizzazione, mettendo chi la dirige o chi la controlla nelle condizioni di prendere migliori decisioni riguardo al suo funzionamento.

Il contenuto dell'articolo è il seguente. La sezione 2 esamina gli aspetti salienti di tre esperienze di analisi di performance condotte negli Stati Uniti. Da queste esperienze generalizziamo, nella sezione 3, alcune indicazioni trasversali, utili, a nostro avviso, per chiunque desideri condurre un'analisi di performance. L'articolo si conclude con brevi riflessioni critiche sull'esperienza italiana.

## 2. **Riflessioni su alcune esperienze statunitensi di *performance measurement***

Negli Stati Uniti esiste una tradizione ultratrentennale di costruzione ed utilizzo di sistemi di “*performance measurement*” al fine di tenere sotto controllo e migliorare la performance di organizzazioni pubbliche. Per quanto si debba essere molto cauti nel trasferire questa esperienza alla realtà italiana, riteniamo che da una sua attenta analisi possano essere tratte alcune utili lezioni di natura generale. In questa sezione passiamo in rassegna tre casi relativi a contesti organizzativi molto diversi tra loro. Piuttosto che illustrare in dettaglio come il sistema di analisi di performance è stato realizzato in questi tre casi, è nostra intenzione far risaltare la filosofia che ne ha guidato il *processo* di costruzione. In particolare ci preme sottolineare come in tutte e tre queste esperienze, nonostante le notevoli differenze che le caratterizzano, la scelta degli indicatori da adottare sia solo un momento intermedio di questo processo. A monte e a valle di questo momento si dedica molta attenzione ad aspetti altrettanto importanti: alla definizione dei bisogni informativi che gli indicatori devono soddisfare; ai possibili modi per consentirne un reale utilizzo all'interno del processo decisionale; ai pericoli di distorsioni dei comportamenti che l'impiego di indicatori può indurre all'interno delle organizzazioni; all'individuazione di adeguati standard e termini di confronto.

Il primo caso si riferisce all'implementazione su larga scala di sistemi di analisi di *performance*, sperimentata all'interno delle maggiori Agenzie federali e promossa da una legge del 1993, il *Government Performance and Results Act*. Nella sezione 2.1 utilizziamo come chiave di lettura di tale esperienza l'interpretazione che ne è stata data dal *General Accounting Office*, la Corte dei conti americana: illustriamo sia la metodologia generale con la quale, secondo il GAO, l'analisi di performance dovrebbe essere impostata e condotta nelle amministrazioni federali, sia gli specifici rilievi critici, da essa sollevati, alla realizzazione di un sistema di programmazione ed analisi di performance da parte della *Federal Emergency Management Agency*, l'Agenzia federale per la protezione civile.

Il secondo caso è relativo alla valutazione della performance di agenzie locali, create da una legge del 1992 (il *Job Training Partnership Act*) ed incaricate di gestire il complesso degli interventi di formazione e aiuto all'inserimento lavorativo per soggetti svantaggiati. La performance di queste agenzie per l'impiego è stata sottoposta a partire dal 1983 ad un sistema di misurazione, con lo scopo di identificare quelle agenzie che operano sistematicamente al di sotto di standard minimi, calcolati tenendo conto delle condizioni del mercato del lavoro in cui ciascuna agenzia si trova ad operare.

Il terzo caso riguarda invece un approccio complessivo al *performance measurement*, formulato da uno dei più noti esperti in questo campo, Harry Hatry, impegnato da trent'anni nella costruzione di metodi semplici e facilmente applicabili per la misurazione della performance di servizi pubblici locali. L'ultimo manuale prodotto da Harry Hatry, intitolato “*Performance Measurement: Getting Results*” contiene una lunga serie di consigli pratici per chi si accinga a costruire un sistema di misurazione della performance di un'organizzazione pubblica. Non è nostro intento riassumere tutti questi consigli pratici, molti dei quali andrebbero “decontestualizzati” per poter essere utilmente applicati alla realtà della pubblica amministrazione italiana. Il nostro obiettivo nel dar conto di questo approccio è più modesto: ci preme metterne in evidenza la filosofia generale e soprattutto l'attenzione che il manuale presta alle fasi del processo che precedono e seguono la costruzione degli indicatori.

## 2.1 Il *Government Performance and Results Act* e il *performance plan* della Protezione Civile

Il *Government Performance and Results Act* (GPRA), approvato nel 1993 dal Congresso degli Stati Uniti, impone ai Ministeri e alle maggiori Agenzie federali di adottare sistemi di misurazione e di analisi delle loro performance. Lo scopo principale è di stimolare tali organizzazioni a rendere conto in modo trasparente allo stesso Congresso e, più in generale, a tutti i cittadini, delle scelte effettuate e dei risultati delle attività di cui sono responsabili (“...*improve the confidence of the American people in the capability of the federal government by systematically holding Federal agencies accountable for achieving program results*”).

In sostanza la legge richiede alle amministrazioni federali di ridefinire, attraverso l’elaborazione di *strategic plans*, le loro missioni istituzionali e di identificare obiettivi generali di lungo periodo che guidino i programmi dell’agenzia per i successivi 5/10 anni. Inoltre esse sono obbligate a presentare con cadenza annuale *performance plans*, documenti programmatici che indichino le attività da svolgere e gli obiettivi da raggiungere nel corso dell’anno successivo, e *performance reports*, relazioni che contengano i risultati delle analisi svolte e spieghino se, e in che misura, gli obiettivi dichiarati sono stati raggiunti, nonché i motivi di eventuali fallimenti e le soluzioni proposte per migliorare prestazioni mediocri o scadenti. Con l’emanazione di tale legge ha avuto quindi inizio un processo di sperimentazione su larga scala, durato circa sei anni<sup>1</sup>; la produzione che ne è scaturita, di rapporti, articoli, commenti e riflessioni sui problemi legati all’implementazione di tali sistemi analitici, è enorme e rappresenta un capitale di conoscenza e di esperienze dal quale è possibile attingere a piene mani.

In seno a tale sperimentazione un ruolo di particolare rilievo è stato svolto dal *General Accounting Office* (GAO), la Corte dei conti americana, che ha accompagnato l’implementazione del GPRA in qualità sia di osservatore critico, per conto del Congresso, sia di punto di riferimento metodologico nei confronti delle Amministrazioni federali. Dal nostro punto di vista appare perciò di grande interesse esaminare i documenti prodotti dal GAO, che rappresentano una testimonianza dello spirito che ha guidato GPRA e delle difficoltà che ne hanno caratterizzato la messa in opera. Essi contengono, in alcuni casi, le raccomandazioni metodologiche generali indirizzate alle Agenzie; in altri, le osservazioni critiche sulle specifiche esperienze condotte all’interno delle singole amministrazioni.

Nel suo impianto metodologico, il GAO (1996) individua tre passaggi chiave che dovrebbero segnare il percorso da compiere verso l’analisi di performance; tali passaggi (*define mission and desired outcome; measure performance; use performance information*) costituiscono il cuore di questo approccio, improntato ad un “sano realismo” riguardo gli eventuali usi a fini decisionali dei risultati dell’analisi e alla consapevolezza dei possibili rischi di distorsioni od errate interpretazioni di tali risultati. Esso soprattutto si fonda sul pieno riconoscimento delle debolezze e dei pericoli insiti nell’applicazione di un metodo d’analisi che ambisce a giudicare il successo di un’organizzazione sulla base di semplici misure quantitative.

### “*DEFINE MISSION AND DESIRED OUTCOMES*”

Le Agenzie federali sottoposte al mandato di GPRA sono organizzazioni di vaste dimensioni, con molti anni di attività e con competenze che insistono su un ambito territoriale molto esteso; esse perciò si trovano spesso nelle condizioni di dover interagire con un gran numero di istituzioni e prestare la loro opera nei confronti di diverse categorie di soggetti (cittadini, imprese, associazioni, enti locali). Per questi motivi le loro attività ben difficilmente possono essere ricondotte al raggiungimento di un unico risultato chiaramente identificabile, ma esse si trovano spesso a svolgere funzioni diverse con finalità assai differenti. In contesti organizzativi caratterizzati da una così elevata complessità, prima di misurare, è necessario ricostruire la missione stessa dell’organizzazione e definire i risultati ai quali essa tende. Ma che cosa significa per un’organizzazione pubblica “*ricostruire la propria missione*”?

In estrema sintesi, significa esplicitare le ragioni per cui quella particolare organizzazione interviene in un determinato settore o ambito di *policy*: quale “problema” di rilevanza pubblica cerca di risolvere od alleviare; a quali fondamentali esigenze della collettività s’impegna a dare risposta; quali soggetti titolari del diritto a

<sup>1</sup> L’attuazione di GPRA è proceduta molto gradualmente: entro il settembre 1997 ogni amministrazione doveva presentare al Congresso il proprio *strategic plan*; prima del gennaio 1999 doveva essere approvato il primo *performance plan*; entro il 31 marzo 2000 doveva essere presentato il primo *performance report*. Negli anni precedenti a queste scadenze molte Agenzie sono state impegnate in esperienze-pilota.

una certa prestazione intende servire. A questo sforzo definitorio dovrebbe naturalmente seguire la specificazione dei risultati che l'organizzazione si prefigge, delle strategie d'azione che ha intenzione di mettere in atto per tentare di raggiungerli e dei motivi per cui si sono scelte tali strategie e non altre possibili.

Affinché questo sia un tentativo serio e non si riduca ad uno sterile esercizio di retorica, il GAO raccomanda ai responsabili dell'organizzazione di seguire alcune "buone prassi di lavoro". In primo luogo, è importante avviare un processo dinamico ed inclusivo, teso a coinvolgere tutti gli *stakeholders*, cioè tutti quei soggetti che hanno un qualche interesse, una qualche posta in gioco (*stake*), nell'area d'intervento dell'organizzazione: dai suoi stessi finanziatori; ai gruppi di interesse che fanno parte del *policy network* di riferimento; ai destinatari/utenti/beneficiari delle politiche che l'organizzazione contribuisce ad implementare; ai rappresentanti di altre amministrazioni. Il significato di tale raccomandazione è chiaro: evitare il rischio che l'amministrazione, troppo concentrata nello sforzo di conoscersi e di guardare al proprio interno, compia un'operazione sostanzialmente "autoreferenziale" e dimentichi di confrontarsi con le istanze provenienti dall'esterno.

In secondo luogo, è necessario tenere nella giusta considerazione i fattori "ambientali", sia esterni che interni all'amministrazione (evoluzione dei fenomeni economici e sociali, cambiamenti nel quadro normativo, preparazione e formazione culturale del personale, livello di informatizzazione interna), che possono incidere sulla capacità dell'organizzazione stessa di raggiungere i risultati desiderati. Una sottovalutazione di tali fattori può portare l'organizzazione a porsi obiettivi troppo ambiziosi e al di fuori della sua portata, o, peggio, a fallire completamente nella scelta delle attività da realizzare. Infine, come ultima raccomandazione, si pone l'accento sulla necessità di mettere in discussione e ripensare le attività tradizionalmente svolte dall'organizzazione, chiedendosi il "se e perché" sia opportuno continuare a svolgere tali attività. L'invito è di sfuggire alla naturale tendenza "autoconservativa" di descrivere ciò che si sta facendo, o che si è abituati a fare, invece di prevedere in modo creativo nuove attività, più rispondenti ai bisogni dei cittadini e agli interessi degli *stakeholders*. Insomma, riassumendo lo spirito delle tre raccomandazioni in uno slogan, le Agenzie sono incoraggiate a "*guardare di più all'esterno delle proprie strutture per capire come organizzarsi ed agire meglio al proprio interno*".

Il caso della *Federal Emergency Management Agency*<sup>2</sup> (FEMA), l'Agenzia federale per la Protezione Civile degli Stati Uniti, rappresenta un buon esempio del modo in cui questo percorso, se affrontato seriamente, possa davvero condurre ad una sostanziale revisione delle finalità e delle strategie dell'organizzazione. Dopo alcune calamità naturali, avvenute tra la fine degli anni ottanta e gli inizi degli anni novanta (l'uragano *Hugo* e il terremoto di *Loma Prieta* nel 1989; l'uragano *Andrew* nel 1992), la reputazione della Protezione Civile era molto appannata agli occhi dell'opinione pubblica statunitense; su di essa gravava un generale scetticismo riguardo la sua capacità di far fronte a situazioni d'emergenza nazionale. GPRA rappresentò allora l'opportunità, per i responsabili della Protezione Civile, di rispondere in modo costruttivo alle critiche e ai dubbi sollevati intorno all'operato dell'Agenzia. Tradizionalmente la FEMA aveva concentrato gli sforzi e le risorse a sua disposizione per fornire assistenza alle vittime di disastri naturali; si interveniva nel momento in cui nasceva lo stato d'emergenza e il fine degli interventi era di alleviare le situazioni di disagio e di sofferenza sorte in seguito all'evento naturale. Durante la stesura del piano strategico, approvato nel 1997 al termine di un ampio processo di riflessione, le modalità d'intervento della FEMA sono state completamente riviste, quasi rovesciate: accanto alle classiche attività *post* disastro, volte all'immediato soccorso e al ripristino delle infrastrutture distrutte, è stato dato grande spazio alle attività, precedentemente trascurate, di prevenzione e di riduzione del rischio per le persone e le proprietà<sup>3</sup>. Piuttosto che darsi obiettivi di miglioramento sulle attività svolte abitualmente, l'organizzazione ha quindi scelto di impegnarsi in programmi innovativi; il processo cognitivo avviato con la ridefinizione della propria missione ha messo in evidenza nuovi bisogni della collettività e ha permesso all'Agenzia di individuare nuovi modi per dar loro risposta.

<sup>2</sup> La *Federal Emergency Management Agency* è stata istituita nel 1979; si tratta dell'agenzia responsabile a livello federale del coordinamento e della programmazione del soccorso e dell'assistenza civile in caso di disastri naturali (alluvioni, inondazioni, uragani e terremoti).

<sup>3</sup> Ad esempio, l'Agenzia ha deciso di finanziare campagne di comunicazione rivolte agli studenti contro il rischio di incendi; di promuovere iniziative che impegnano gli Stati, le Autorità locali e i rappresentanti delle imprese costruttrici ad adottare codici di sicurezza per la costruzione di infrastrutture più resistenti alle calamità naturali; di formare agenti di polizia specializzati nel recupero di individui isolati o dispersi durante alluvioni o inondazioni.

## “MEASURE PERFORMANCE”

Il secondo passaggio che il GAO individua è relativo alla misurazione della performance, al tentativo cioè di dare una rappresentazione quantitativa del successo ottenuto dall'organizzazione nel raggiungere una determinata finalità. In questa sede però non vengono proposte tipologie o batterie di indicatori predefiniti da adottare/adattare nei diversi contesti organizzativi; si preferisce piuttosto indicare quali sono le cautele che un valutatore dovrebbe seguire nella scelta degli indicatori. Non vengono offerte facili scorciatoie. Al contrario; i criteri che il GAO definisce complicano il lavoro del valutatore e lo costringono, prima di accettare un indicatore come una corretta misura di successo, a testarne ripetutamente la validità e a verificare le possibilità di suo effettivo utilizzo.

Il primo compito del valutatore consiste nell'esplicitare chiaramente in che modo l'attività svolta dall'organizzazione possa influire sui valori assunti dall'indicatore; se non esiste un chiaro e dimostrabile legame tra ciò che l'organizzazione compie e ciò che l'indicatore misura, è opportuno rifiutarne l'impiego. Anche se tale connessione esiste, occorre comunque fare lo sforzo, prima di adottare l'indicatore come misura di performance, di individuare e chiarire quali sono i fattori esogeni, indipendenti dall'operato dell'organizzazione, che ne possano influenzare il valore; dall'adeguato riconoscimento di tali fattori dipende infatti la possibilità di utilizzare l'indicatore a scopo valutativo.

In secondo luogo, accanto all'indicatore deve essere anche individuata l'unità organizzativa responsabile di un suo eventuale miglioramento, o peggioramento, che possa/debba reagire, con azioni opportune, ai segnali che provengono da tali variazioni. Senza questo sforzo di attribuzione di responsabilità è improbabile che i valori assunti dall'indicatore possano essere d'aiuto nell'orientare future decisioni. In terzo luogo, il valutatore deve resistere alla tentazione di usare “troppi indicatori”. Spesso, infatti, si tenta di sopperire all'intrinseca limitatezza informativa di questo strumento, aumentando artificialmente il numero degli indicatori individuati; questa operazione però rischia di complicare inutilmente la lettura dei dati e di rendere ancora più difficile interpretarne il senso. Infine, ogni attività di misurazione nasconde delle insidie nel momento in cui si passa alla raccolta delle informazioni. Questa è una fase che viene spesso trascurata e di cui si sottovaluta l'importanza; in realtà dietro ad ogni indicatore si possono nascondere problemi di attendibilità, accuratezza e completezza dei dati, causati dalle modalità di raccolta e di rilevazione delle informazioni. Non considerare fin dall'inizio tali difficoltà significa far perdere di credibilità all'intero processo di misurazione.

Le obiezioni sollevate dal GAO ai tre *performance plans* fino ad oggi presentati dalla Protezione Civile si ispirano direttamente ai criteri appena esposti e ne rappresentano una concreta applicazione. La disamina del GAO ha l'intento di verificare se gli indicatori adottati consentiranno davvero al Congresso di capire in che direzione si stia muovendo l'Agenzia, in che misura essa stia raggiungendo gli obiettivi fissati nello *strategic plan* e in che modo si stia attrezzando per rispondere a eventuali performance negative. Ogni sottovalutazione da parte dell'Agenzia dei limiti del sistema d'analisi proposto viene quindi opportunamente evidenziata e sottoposta a critica; qualche esempio dei rilievi mossi ci permette di comprenderne meglio lo spirito:

- FEMA propone circa 150 indicatori di performance: molti tra questi sono però ridondanti e non hanno nessun reale valore conoscitivo per il Congresso, il GAO ne consiglia perciò l'eliminazione dal *performance plan*;
- accanto ad ogni indicatore è prevista una sezione dedicata a “*mezzi e strategie*” con cui s'intende operare su quella misura al fine di migliorarne il valore: in molti casi però non è chiaro quali sono le strategie che l'Agenzia intende adottare o manca un nesso evidente tra la strategia adottata ed un eventuale cambiamento nel valore della misura;
- in molti casi si ignora che esistono fattori esterni (ad esempio l'attività di altre organizzazioni) che possono influenzare in modo determinante il valore assunto dall'indicatore; per tutti questi casi l'Agenzia dovrebbe specificare se e in che modo prevede di tenere sotto controllo l'incidenza di tali fattori;
- accanto ad ogni indicatore è prevista una sezione dedicata alla “*verifica della validità*” dell'informazione raccolta: spesso l'Agenzia mostra superficialità nel trattare questo problema, non mette in evidenza i

limiti delle proprie fonti informative e non considera la possibilità di usare informazioni alternative per convalidare i risultati ai quali è pervenuta.

*“USE PERFORMANCE INFORMATION”*

La definizione della missione, la scelta degli indicatori e la raccolta delle informazioni non esauriscono il percorso analitico da intraprendere al fine di dare un giudizio sulla performance di un'organizzazione. E' necessario effettuare un ultimo sforzo: rendere le informazioni utilizzabili ed utili al processo decisionale. Non porre sufficiente attenzione a quest'ultima fase dell'analisi significa svuotare di valore e di sostanza anche le due fasi precedenti.

In particolare si sottolinea l'importanza di avere degli standard di riferimento e la necessità di rendere chiaramente interpretabili, anche ad interlocutori esterni all'organizzazione, le differenze riscontrate tra questi e i valori realmente assunti dagli indicatori. Nel caso di GPRA tali standard assumono di solito la forma di obiettivi di miglioramento rispetto alla performance realizzata nell'anno precedente o comunque rispetto una *baseline*; non vengono quasi mai proposti confronti con altre organizzazioni. Ciò è dovuto principalmente al carattere di unicità delle Agenzie federali, nonché delle attività che esse svolgono sul territorio nazionale, per le quali sarebbe molto difficile individuare dei validi termini di confronto al di fuori della stessa amministrazione.

Anche a questo proposito le critiche effettuate dal *General Accounting Office* ai *performance plans* della Protezione Civile sono chiarificanti; il GAO, infatti, non si accontenta che all'interno del piano siano individuati degli indicatori di performance, in qualche modo connessi al successo dell'organizzazione, ma pretende che ai valori eventualmente assunti da tali indicatori sia data un'interpretazione univoca. Se, ad esempio, si prevede di utilizzare quale indicatore di performance il *“tempo necessario ai soccorsi per giungere sul luogo del disastro dal primo segnale di aiuto ricevuto”*, occorre anche specificare chiaramente rispetto a quale indice verrà espresso il giudizio (tempo medio, mediano, massimo) e perché; quale miglioramento si prevede di raggiungere su tale indice e come è stato determinato tale obiettivo; rispetto a quale periodo di riferimento (l'anno passato, la media dei tre anni precedenti); quali potrebbero essere i motivi di un eventuale fallimento nel tentativo di raggiungere tale obiettivo. Soltanto una chiara specificazione e comunicazione<sup>4</sup> di tali aspetti rafforzano le probabilità che tali misure saranno effettivamente usate a fini decisionali e che diventino effettivi strumenti di giudizio sull'operato dell'organizzazione.

Per concludere, lo sforzo analitico compiuto dalle Agenzie federali e dal GAO, che abbiamo messo in evidenza in queste pagine, dovrebbe soprattutto consegnarci un'importante lezione: sistemi di misurazione di questo tipo, concepiti a fini rendicontativi, sono davvero utili solo se vi è nei soggetti interessati una piena attenzione alla rilevanza, validità e comunicabilità delle informazioni raccolte. Nel caso contrario vi è il rischio concreto che tutta l'operazione si trasformi in un mero esercizio contabile, magari molto costoso ed impegnativo, ma sicuramente poco utile a prendere decisioni.

## **2.2 Il Job Training Partnership Act e i performance standards per le agenzie locali per l'impiego**

Questa esperienza di analisi della performance si differenzia da quella precedente sotto vari profili: innanzitutto riguarda un solo tipo di intervento pubblico, la fornitura di servizi formativi e di orientamento per soggetti economicamente svantaggiati, invece che applicarsi a tutto lo spettro delle politiche del governo federale. Una legge emanata nel 1982, il *Job Training Partnership Act*, rivoluziona il modo con cui si finanziano e si erogano servizi per favorire il reinserimento lavorativo delle fasce deboli. Vengono istituite nuove agenzie per l'impiego a livello locale, le *Service Delivery Areas*, con il compito di organizzare l'offerta dei servizi sul territorio, gestire la selezione e l'orientamento degli allievi/utenti, appaltare lo svolgimento dei corsi ai centri di formazione e ad altre organizzazioni pubbliche e private, raccogliere i dati sulle caratteristiche degli allievi e sui loro esiti occupazionali. Esistono circa 600 agenzie locali per l'impiego in tutti gli Stati Uniti: esse dipendono gerarchicamente dall'amministrazione di ciascun Stato, che

<sup>4</sup> Donald Kettl (citato in Newcomer, 1997) sottolinea che uno dei maggiori benefici derivati dall'implementazione di sistemi di misurazione della performance nel governo federale è stato proprio il miglioramento della comunicazione dei programmi e degli obiettivi delle agenzie a favore di interlocutori esterni quali il Congresso.

ha compiti di supervisione su di esse. L'opera di supervisione da parte dell'amministrazione di ciascun Stato si basa in parte sull'uso dei *performance standards* descritti in questa sezione.

I servizi offerti ai disoccupati dalle agenzie locali variano lungo un ampio spettro che va dall'offerta di corsi di formazione tradizionale (*classroom training*) a forme di apprendistato/formazione sul lavoro (*on-the-job training*) a corsi per imparare a cercare lavoro (*job search assistance*). I servizi non sono prodotti direttamente dall'agenzia locale per l'impiego, bensì appaltati a organismi pubblici e non-profit, quali i centri di formazione, i *community colleges*, gli istituti medi superiori, le organizzazioni filantropiche e del volontariato.

Una novità importante che JTPA introduce è l'istituzione di un sistema di misurazione della performance, che intende valutare la capacità delle agenzie di ottenere buoni esiti occupazionali per gli utenti dei servizi. Poiché il recupero dello svantaggio economico, cioè dello stato di povertà, è l'*obiettivo di fondo* di JTPA, è naturale che si usi il miglioramento delle prospettive occupazionali degli allievi come criterio-guida per giudicare la performance delle agenzie locali. Il primo passo della costruzione di questo sistema è quindi quello di identificare con precisione le *dimensioni di successo* sulla base delle quali l'attività svolta dalle agenzie va giudicata. Le principali dimensioni di successo vengono identificate nella probabilità di trovare lavoro tra gli utenti dei servizi entro un certo periodo dalla frequenza/fruizione dei servizi; e nella retribuzione ottenuta dagli allievi che hanno trovato lavoro.

Queste dimensioni di successo vengono poi tradotte operativamente in *indicatori di performance*. Sono stati utilizzati solo due indicatori, replicati però per diverse combinazioni di età e sesso degli allievi:

- la *percentuale di occupati* (per almeno 20 ore alla settimana) tra gli allievi, rilevato mediante un indagine di follow-up a 13 settimane dalla conclusione del corso/fruizione del servizio;
- la *retribuzione settimanale*, rilevata per coloro che sono occupati alla 13-esima settimana dalla conclusione del corso/fruizione del servizio.

L'adozione di questi indicatori implica delle scelte metodologiche molto precise. Innanzitutto implica la volontà di porre tutta l'enfasi sui risultati finali (*outcomes*) ottenuti dai servizi, a scapito degli indicatori di attività (ad esempio: il numero di ore di lezione effettuate) o di *output* (ad esempio: il numero di utenti serviti) o di soddisfazione della domanda (il rapporto tra numero di utenti serviti e il numero di utenti potenziali) o di costo (costo per utente servito.) La scelta è quindi quella di privilegiare, nel giudicare la performance di questo servizio, i risultati ottenuti a favore degli utenti.

La seconda scelta "forte" che discende dall'adozione di questi due indicatori è quella di focalizzare l'attenzione sui soli risultati legati al successo sul mercato del lavoro. La frequenza di corsi di formazione o la fruizione di altri servizi per l'impiego può produrre a favore degli utenti "effetti" positivi su altre dimensioni, quali l'autostima e la capacità di socializzazione, che non si traducono direttamente in un maggiore successo nel trovare lavoro o in una maggiore retribuzione. Tuttavia, queste dimensioni di successo sono molto più difficili da misurare e comunque sono state ritenute secondarie ai fini di dare un giudizio sulla performance di queste agenzie.

A questo punto è importante evidenziare i due *passi ulteriori* che vengono compiuti nella costruzione di questo sistema di misurazione della performance: (i) la definizione, rispetto a ciascun indicatore, di un *livello minimo di performance* che ogni agenzia è tenuta a raggiungere in un certo anno; e (ii) la creazione di un *meccanismo di sanzioni* per quelle agenzie la cui performance si riveli sistematicamente (cioè per tre anni consecutivi) ad di sotto dello standard.

La parte più interessante e innovativa di questo sistema riguarda appunto la *fissazione dello standard* per ciascun indicatore. Non ci si è limitati a definire un valore arbitrario comune a tutte le agenzie, bensì si è costruito un meccanismo per creare, ogni anno, uno standard "calibrato" per ciascuna delle 600 agenzie. Lo scopo della calibrazione è quello di assegnare a ciascuna agenzia un valore che sia tale da: (i) non penalizzare le agenzie che operano in condizioni socio-economiche più difficili di altre, e quindi hanno più difficoltà a garantire buoni esiti occupazionali; e (ii) non penalizzare quelle agenzie che servono una popolazione di utenti più "difficile", cioè con caratteristiche meno appetibili ai datori di lavoro.

Questa seconda esigenza va sottolineata. L'intento è prevenire un problema molto comune nei servizi mirati a soggetti portatori di problematiche particolari: il pericolo che il servizio operi una *scrematura*<sup>5</sup> dei propri utenti, in modo da evitare i casi più difficili, in questo caso quelli meno occupabili. L'esistenza di un sistema di misurazione della performance può paradossalmente acuire questo pericolo: se le agenzie sono giudicate sulla base degli esiti occupazionali che i loro utenti raggiungono, hanno l'incentivo a mettere in atto tattiche di scoraggiamento nei confronti degli utenti più difficili da occupare, in modo da poter tenere alta la propria performance. Questo pericolo è acuito nel caso in cui la loro performance sia giudicata sulla base di uno standard unico, che non tiene conto delle diverse condizioni locali e della composizione della platea degli utenti. Per garantire uguali condizioni di giudizio, occorre quindi assegnare uno standard proporzionalmente più basso a quelle agenzie che operano in una zona con più alto tasso di disoccupazione o con retribuzioni medie più basse e/o che attraggono soggetti con caratteristiche personali che costituiscono fattori di svantaggio sul mercato del lavoro (ad esempio, l'essere anziano, nero, disabile, disoccupato da molto tempo, senza dimora, analfabeta, ecc.).

Questo risultato è stato ottenuto mediante il ricorso ad un modello statistico (regressione) che identifica qual è l'effetto marginale di ciascun *fattore di svantaggio* sul valore medio dell'indicatore (sia il tasso di occupazione sia la retribuzione media) a livello nazionale. Il compito di stimare il modello viene affidato ad un istituto di ricerca che opera su contratto con il Ministero del Lavoro e che ha accesso ai dati su tutte le 600 agenzie. Ogni anno questo istituto stima una nuova versione del modello e genera un insieme di "pesi", uno per ciascun fattore di svantaggio. Moltiplicando il peso per il corrispondente fattore di svantaggio osservato in ciascuna agenzia, si ottiene un valore di correzione da applicare a ciascuna agenzia per ottenere il valore atteso della performance dell'agenzia, e quindi il suo standard di performance.

I dettagli di questa procedura sono illustrati in Barnow (1992) e in Dickinson (1988). In questa sede non interessa tanto entrare in ulteriori dettagli operativi, quanto passare in rassegna gli ingredienti essenziali di questo approccio:

- si fa innanzitutto chiarezza su quelli che sono gli obiettivi dell'azione, e da questi si ricavano le dimensioni di successo sulla cui base giudicare la performance delle agenzie;
- dalle dimensioni di successo si ricavano gli indicatori, che vengono concretamente calcolati ogni anno per ciascuna agenzia, ricavando i dati da un'indagine di follow-up;
- si definisce uno standard con cui confrontare la performance osservata ogni anno per ciascuna agenzia;
- ci si cautela contro i possibili effetti distorsivi che l'uso di un unico standard per tutte le agenzie potrebbe avere, in particolare sulla possibilità di scrematura a danno degli utenti meno facilmente occupabili;
- si mette in piedi un sistema statistico che identifica, per ciascuna agenzia, un livello atteso di performance calibrato sulle caratteristiche dell'economia locale e sul mix di caratteristiche socio-economiche dei suoi utenti;
- si introduce un meccanismo di sanzioni per quelle agenzie la cui performance è sistematicamente al di sotto dello standard così calcolato.

Si tratta quindi di un sistema di misurazione della performance che coniuga rigore metodologico e attenzione per l'utilizzo concreto delle informazioni prodotte, una combinazione abbastanza rara anche nel panorama statunitense.

### 2.3 L'approccio di "*Performance Measurement: Getting Results*"

Il contenuto del manuale "*Performance Measurement: Getting Results*" (Hatry, 1999) è frutto di una pluridecennale esperienza da parte del suo autore e di un gruppo di analisti a lui associati presso l'Urban Institute e presso l'International City/County Management Association (ICMA), l'organizzazione che rappresenta i manager degli enti locali (città e contee) statunitensi. Il manuale contiene una grande quantità di spunti e consigli pratici per mettere in piedi un sistema che misuri la performance di un *program*, cioè di

<sup>5</sup> Nella letteratura anglosassone questo fenomeno è generalmente chiamato "*creaming*" (Barnow, 2000).

un programma di intervento pubblico. Il *program*, dotato di un suo budget e diretto al raggiungimento di uno scopo relativamente ben definito, è il modo fondamentale di operare del settore pubblico negli Stati Uniti. Tutta la valutazione, compresa la misurazione della performance, tende a ruotare attorno a questo concetto, che peraltro è difficilmente traducibile nel contesto italiano, dove è più importante l'enfasi sulla struttura organizzativa a livello territoriale.

I campi principali di applicazione dell'approccio descritto dal manuale sono i servizi pubblici erogati da enti del governo locale, direttamente o per il tramite di organizzazioni *non-profit*. Più raramente le applicazioni riguardano interventi erogati dal governo federale: a differenza di GPRA, illustrato nella sezione 2.1, questo approccio non riguarda la performance di grandi e complesse organizzazioni, quali agenzie federali, ma singoli programmi di intervento messi in atto da singole organizzazioni.

Quella che presentiamo in questa sezione non è una dettagliata recensione del manuale, ma un tentativo di metterne in luce gli elementi distintivi che a nostro parere lo rendono rilevante anche al di là del contesto statunitense a cui più direttamente si applica. Sono questi elementi di impostazione generale da cui è possibile trarre qualche lezione utile, mentre la mole di consigli pratici che il manuale offre è di più difficile fruibilità e applicabilità, perché molto legati al particolare contesto americano. Secondo questa lettura, abbiamo individuato quattro "elementi" che a nostro parere rendono questo approccio rilevante al di là del contesto specifico: (i) l'idea di fondo che l'analisi della performance non si riduca ad un'attività di misurazione, ma sia invece un processo complesso che serve in ultima istanza a "dare significato" alle misurazioni effettuate; (ii) la necessità, come primo passo di questo processo, di chiarire la missione e gli obiettivi di miglioramento perseguiti dall'organizzazione; (iii) l'importanza di affinare la fase di definizione e costruzione degli indicatori; (iv) l'attenzione dedicata a far sì che il prodotto della misurazione venga effettivamente utilizzato.

*"THINK BEFORE YOU MEASURE ..."*

Va messo in evidenza come il manuale non si apra dando prescrizioni su cosa sia necessario misurare né su quali indicatori sia necessario usare, né presentando sistemi per classificare a priori gli indicatori. Anzi, la costruzione degli indicatori non viene trattata in maniera esplicita fino quasi alla metà del manuale. L'enfasi non è quindi, fin dall'inizio, sugli aspetti della misurazione, bensì sul *processo* che è importante attivare per dare un senso alla misurazione e facilitarne l'utilizzo. La parte centrale del manuale si intitola appunto "*The performance measurement process*" e contiene una descrizione dettagliata dei passi necessari a mettere in piedi un processo di questo tipo.

Il manuale mette in evidenza i prerequisiti per avviare questo processo: (a) avere l'appoggio attivo da parte del top management dell'organizzazione; (b) operare in un contesto stabile: servizi che stanno attraversando fasi di ristrutturazione o ridefinizione di ruoli non sono buoni candidati per una misurazione della performance; (c) avere a disposizione un minimo di "*computerized data-processing capability*": quindi non di dati già disponibili, ma della capacità di processare dati mediante computer. Questi tre requisiti, pur nella loro semplicità (e volendo banalità), sottolineano tre fatti importanti: se gli utilizzatori finali non sono convinti e coinvolti, non conviene neppure partire con l'attivazione del processo; un sistema di misurazione standardizzata richiede stabilità, senza la quale la standardizzazione non è possibile; i "dati" non devono necessariamente già esistere (e la loro mancanza non è una scusa sufficiente per non attivare il processo), ma devono esserci le condizioni minime per generarli.

II. "*WHAT ARE THE PROGRAM'S MISSION AND OBJECTIVES?*"

Molti dei capitoli del manuale sono formulati sotto forma di domanda a cui il valutatore deve rispondere per compiere i passi successivi del processo di costruzione del sistema. La prima importante domanda è: "Qual è la missione e quali sono gli obiettivi che il servizio (o l'organizzazione) sta perseguendo?" Un elemento va in particolare sottolineato riguardo alla discussione contenuta in questa parte del manuale: non si presuppone che la definizione di missione ed obiettivi esista già, in quanto generata esplicitamente dal processo decisionale, bensì si accetta il fatto che la missione venga definita come parte del processo stesso di analisi di performance. Questo atteggiamento a nostro parere contrasta con quello riassunto nella frase "non si può valutare finché il processo decisionale non ha chiaramente e univocamente stabilito quali sono gli obiettivi". E' difficile incontrare un processo decisionale che univocamente fissi missioni e obiettivi da

raggiungere: questi, nella maggior parte dei casi, vengono enunciati in modo molto generico o mancano del tutto.

Il manuale quindi riconosce che l'erogazione di servizi pubblici non segue un modello razional-sinottico in cui prima vengono fissati gli obiettivi e poi deduttivamente si procede a stabilire come raggiungerli. La missione dell'organizzazione e i suoi obiettivi vanno ricostruiti in modo pragmatico nel momento in cui ci si pone il compito di valutare la performance. A questo scopo il manuale dà una serie di consigli pratici e lista una serie di potenziali fonti informative da cui trarre ispirazione per questa opera di ricostruzione: al primo posto stanno leggi e regolamenti, seguiti dai bilanci, piani strategici, discussioni con il top management, incontri con operatori e utenti, contatti con altri enti che offrono servizi analoghi. Ciò che è importante sottolineare è che, in questa prima fase, si rimane ancora lontani dalla preoccupazione di definire "cosa c'è da misurare", concentrandosi invece sul "cosa definisce il "successo" dell'organizzazione nel perseguimento dei suoi scopi".

Un passo di avvicinamento verso la misurazione è riassunto dalla domanda successiva "*What outcomes should be tracked?*", cioè quali risultati desiderabili devono essere tenuti sotto osservazione? Il manuale distingue quindi tra "*objectives*" e "*outcomes*", una distinzione non facile da rendere in italiano. Secondo il manuale, *objectives* sono gli obiettivi espressi ancora in termini generali e qualitativi: ad esempio "favorire l'inserimento lavorativo dei portatori di handicap", "aiutare le piccole aziende ad esportare", "migliorare la circolazione degli autobus urbani". *Outcomes* sono invece i risultati attesi, in termini di cambiamento/miglioramento osservabile: ad esempio, "aumento dell'inserimento lavorativo dei portatori di handicap", "incremento nelle esportazioni delle piccole imprese", "miglioramento della velocità commerciale degli autobus". I risultati attesi vanno tendenzialmente concettualizzati in termini quantitativi e sono un ulteriore passo verso la definizione degli indicatori.

### III. "*WHAT OUTCOME INDICATORS SHOULD BE TRACKED?*"

La domanda "quali indicatori è più opportuno costruire" viene posta solo a questo punto, presupponendo quindi che il processo sia arrivato a chiarire quali obiettivi il servizio persegue e quali risultati ci si attende dalla sua erogazione. Il manuale mette in chiaro che "*outcome indicators are not the same as outcomes*": gli indicatori di risultato non sono la stessa cosa dei risultati attesi. Gli indicatori sono solo strumenti con cui si misura qualcosa che si ritiene importante e che ci aspetteremmo di ottenere dall'erogazione del servizio. Ad ogni risultato atteso possono corrispondere più indicatori. Il manuale usa l'esempio classico della raccolta rifiuti: accanto a quattro obiettivi generali del servizio in cui si articola la sua missione (migliorare l'estetica; migliorare le condizioni igienico-sanitarie; minimizzare gli inconvenienti per i cittadini; massimizzare la soddisfazione degli utenti), elenca nove risultati attesi (tra cui: grado di pulizia delle strade; assenza di odori; assenza di rumori molesti; perdita di rifiuti durante la raccolta; danni materiali causati durante la raccolta) e quindi 16 indicatori. Tutti gli indicatori elencati assumono la forma di semplici quantità (ad esempio "numero di veicoli abbandonati") o più spesso di rapporti percentuali ("percentuale di famiglie intervistate che si lamenta di rumori molesti"; "percentuale di famiglie che denuncia danni alla proprietà").

La discussione sulle fonti di dati conclude la parte del manuale che si riferisce alla messa a punto del sistema di costruzione degli indicatori. A questo punto il manuale potrebbe interrompersi: consigli pratici sono stati forniti su come costruire gli indicatori e su come raccogliere i dati necessari per tradurli in concrete misurazioni. Il manuale invece non si interrompe, anzi contiene ancora cinque capitoli, dedicati all'*utilizzo* degli indicatori. In questo riteniamo il manuale si stacchi più nettamente da quella parte della pubblicistica italiana sull'argomento che tende a focalizzare l'attenzione esclusivamente sulla costruzione degli indicatori.

### IV. "*MAKING OUTCOME INFORMATION USEFUL*"

L'ottavo capitolo del manuale si apre con una frase che sottolinea bene l'essenza del problema: "*Producing data does not mean that they will be useful*", produrre dati non significa che saranno utili. L'idea di fondo di questo e del successivo capitolo è la seguente: l'informazione contenuta negli indicatori di performance può essere utile solo se si trovano modi per *confrontare* i valori generati dagli indicatori con altri valori, in modo da poter far emergere dal confronto gli elementi per un giudizio. In altre parole, una volta raccolti i dati per "riempire" gli indicatori, i valori ottenuti daranno un'informazione veramente utile solo nella misura

in cui siano confrontati con “qualcosa” e dal confronto emerga la possibilità di dare un qualche giudizio sulla performance osservata.

Il manuale introduce la distinzione tra due famiglie o tipologie di confronti. Una è quella della semplice disaggregazione interna degli indicatori lungo dimensioni osservabili: per definire questo tipo di confronti il manuale usa il termine di *indicator breakout* (disaggregazioni degli indicatori). L'altra è quella dei confronti con valori “esterni”, per il quale il manuale utilizza il termine di *benchmarks*.

I confronti effettuati mediante gli *indicator breakout* sono di natura esplorativa, tendono cioè a stimolare delle domande più che a consentire conclusioni o giudizi. Servono a sollevare questioni del tipo: una performance migliore viene osservata in alcuni casi e non in altri? Esistono delle disparità di trattamento tra categorie di utenti? Diverse modalità di prestazioni corrispondono a risultati diversi? Il manuale fornisce una gran quantità di consigli pratici su come disaggregare gli indicatori raccolti da una singola organizzazione in un certo periodo di tempo: si tratta di confronti relativamente facili da condurre, purché la loro necessità sia riconosciuta e anticipata nella fase di raccolta dei dati. Esempi di questi confronti sono quelli tra sub-unità interne all'organizzazione; tra singoli progetti in cui si articola il programma di intervento; tra categorie di utenti; tra circoscrizioni amministrative; tra diverse tipologie di prestazioni offerte. Il manuale non offre regole generali e universalmente valide per effettuare ogni tipo di confronti, ma riesce ciononostante a mandare un messaggio molto chiaro: il valore informativo degli indicatori ottenuti può venire aumentato a partire da semplici e apparentemente banali operazioni di disaggregazione e confronto interno all'organizzazione.

Il capitolo successivo passa a discutere il secondo tipo di confronti, quelli con i “*benchmarks*” esterni. Il manuale lista sei tipologie diverse di *benchmarks*, e fornisce per ciascuna di esse una discussione di come e quando un tale tipo di confronto può generare informazione utile: (a) confronti con la performance passata per la stessa organizzazione; (b) confronti con la performance di organizzazioni simili; (c) confronti con uno standard generalmente accettato; (d) confronti con organizzazioni del settore privato che producono lo stesso tipo di prestazioni; (e) confronti con modalità diverse di erogare lo stesso servizio; (f) confronti con *target*, cioè obiettivi numerici, stabiliti all'inizio del periodo di performance. Il manuale offre cioè un'interpretazione molto ampia al termine *benchmark*, che va quindi tradotto come “termine di paragone”.

La necessità di esplicitare dei *benchmark* come punto di arrivo del processo di misurazione della performance è chiaramente e semplicemente affermata all'inizio del nono capitolo “*Once a program has outcome data, how can it find out whether the level of performance is good or bad? By comparing the outcomes of the period of interest with benchmarks—that is, measures of what can be expected.*”<sup>6</sup> In altre parole, secondo il manuale, una valutazione, cioè un giudizio, può solo venire dal confronto dei valori degli indicatori ottenuti per quella organizzazione e per quel periodo con altri valori che rappresentano un termine di paragone. Il significato e l'interpretazione del confronto varierà a seconda di come il termine di paragone viene scelto. Non vi sono regole generali, dipende dal contesto specifico e dallo scopo della valutazione quali *benchmark* utilizzare. Quello che il manuale fa egregiamente è mettere l'accento su quello che a nostro parere è il cuore di ogni analisi di performance: preoccuparsi di stabilire quali confronti effettuare con gli indicatori che si sono costruiti.

### **3. Alcuni criteri generali per impostare un'analisi di performance**

I tre casi tratti dall'esperienza statunitense mettono in evidenza come non esista una metodologia unica per la costruzione e l'utilizzo di indicatori di performance. Tuttavia, esistono consapevolezze “trasversali” che tendono a ricorrere, seppure con intensità diverse, in tutti e tre i casi. Lo scopo di questa sezione è mettere a fuoco queste consapevolezze, che, a nostro parere, sono utili a chiunque si accinga a valutare la performance di un'organizzazione attraverso indicatori. Non intendiamo quindi proporre una metodologia applicabile ad ogni contesto; il nostro scopo è di identificare un minimo comun denominatore di cautele da rispettare nei

---

<sup>6</sup> “Una volta che un programma di intervento ha a disposizione dati sui risultati ottenuti, come fa a determinare se essi rappresentano un livello di performance buono o cattivo? Solo confrontando i risultati ottenuti nel periodo di interesse con dei termini di paragone - più precisamente con misure di ciò che ci si attende” (Hatry, 1999)

casi in cui il valutatore abbia l'ambizione di “*dare un giudizio su come funziona un'organizzazione pubblica utilizzando semplici operazioni di misurazione della sua performance, cioè gli indicatori*”.

Abbiamo identificato tre passaggi fondamentali che a nostro parere è necessario affrontare per impostare una proficua analisi di *performance*. Tali passaggi non devono essere interpretati come tre fasi temporalmente successive, ma piuttosto come tre momenti di un unico ragionamento complessivo, separati più da un punto di vista logico che cronologico<sup>7</sup>. Per sottolineare questo concetto, abbiamo formulato questi passaggi sotto forma di domande, a cui il valutatore dovrebbe rispondere non necessariamente nell'ordine in cui sono presentate:

- **Perché** misurare la performance di un'organizzazione: chi può utilizzare questa informazione, con quale potenziale ricaduta su quale segmento del processo decisionale?
- **Cosa** si vuole e può misurare: quali sono le dimensioni di successo dell'organizzazione e con quali indicatori possono essere quantificate?
- **Come** si può interpretare ciò che si è misurato: quali confronti occorre fare per poter esprimere un giudizio sulla performance dell'organizzazione?

### 3.1. *Perché misurare la performance di un'organizzazione?*

La prima domanda che dovrebbe porsi chi si accinga ad elaborare indicatori di performance è la seguente: *Quali sono le fondamentali esigenze informative a cui l'analisi intende dare risposta? E soprattutto: quali soggetti potranno reagire alle informazioni prodotte e attraverso quale meccanismo potrebbero venire indotti a modificare i loro comportamenti/decisioni?*

Una prima importante distinzione è quella tra destinatari interni e destinatari esterni all'organizzazione. La misurazione della performance può essere uno strumento di controllo direzionale interno ad un'organizzazione oppure, cosa ben diversa, può essere uno strumento per rendere conto all'esterno della performance ottenuta. Nel primo caso le misurazioni effettuate devono servire a *segnalare a chi sta al vertice* se l'organizzazione o parti importanti di essa si muovono nella direzione desiderata (in termini di costi, di qualità delle prestazioni, di soddisfazione della domanda di servizi) in modo da poter intervenire con misure correttive. Nel secondo caso le misurazioni effettuate servono invece a *comunicare a soggetti esterni* all'organizzazione informazioni relative alle stesse dimensioni. In questo secondo caso l'esigenza a cui l'analisi di performance intende rispondere è tipicamente quella di *accountability*, cioè il dovere di render conto ad altri dei risultati delle proprie azioni.

Per illustrare questa distinzione conviene ritornare ai tre casi tratti dall'esperienza statunitense. La misurazione della performance prevista da GPRA è un caso che soddisfa essenzialmente un'esigenza di *accountability*: tutto l'apparato di indicatori, obiettivi, *performance* e *strategic plans* serve a mettere le varie amministrazioni federali in grado di rendere conto al Congresso, e quindi indirettamente ai cittadini, di come le risorse pubbliche a loro affidate vengono utilizzate e con quali risultati. La strumentazione descritta nella sezione 2.1 è chiaramente inadatta, per la sua natura complessa e il livello molto aggregato delle misurazioni, ad essere utilizzata per scopi di controllo direzionale all'interno delle organizzazioni. Seppure su un livello diverso, anche la misurazione della performance delle agenzie di JTPA risponde ad una logica di *accountability*: mediante gli indicatori e il confronto con gli standard ogni anno ciascuna agenzia locale è chiamata a rispondere del proprio operato di fronte al livello di governo superiore, che può irrogare sanzioni o premiare. E' quindi difficile pensare che il management di ogni agenzia possa utilizzare quegli stessi indicatori per finalità di controllo dell'attività interna all'agenzia.

Diversamente dai primi due casi, l'ispirazione di fondo del manuale di Hatry è quella di fornire uno strumento semplice ed agile per dirigere meglio un'organizzazione che eroga servizi. Per quanto l'approccio di Hatry tenda ad essere multiuso, raramente si incontra nel manuale la logica dell'*accountability*, mentre molto più spesso si trova l'idea che questo sia uno strumento che possono usare i *public managers* per migliorare il funzionamento della propria organizzazione. Il manuale lo afferma con chiarezza nell'apertura

<sup>7</sup> Kathryn Newcomer (1997) parla di vere e proprie sfide nelle quali l'analista di *performance* si deve cimentare se intende fornire, attraverso la misurazione, informazioni utili per chi decide.

del decimo capitolo: “ *Using performance information only for accountability to higher level would be a great waste. Performance measurement data should be used to help improve programs.*”<sup>8</sup>

Al di là di queste tre esperienze, la distinzione tra controllo direzionale e *accountability* ha implicazioni generali per come si imposta l’analisi della performance, e quindi anche per come si scelgono e usano gli indicatori. Un utilizzo a fini di controllo direzionale sposta l’attenzione su strumenti agili, adattabili alla specificità di ogni servizio, soprattutto con una cadenza dettata dai tempi di erogazione del servizio e di correzione da parte del management. Alcuni indicatori possono essere annuali, ma altri vanno prodotti mensilmente o settimanalmente. In un’ottica di controllo direzionale, i confronti saranno più spesso condotti tra il valore ottenuto dagli indicatori e target prestabiliti o livelli di performance ottenuti in periodi precedenti. Quando invece l’analisi di performance deve servire a render conto all’esterno, si privilegiano tempi più lunghi, indicatori meno calibrati sulla specificità dei servizi, e una tempistica determinata più dalle cadenze della rendicontazione (annuali, pluriennali) che dalla realtà del servizio. I confronti saranno più spesso con *benchmarks* esterni, che consentono meglio a chi sta al di fuori di apprezzare come l’organizzazione si posiziona rispetto ad organizzazioni analoghe.

Possono essere fatte altre distinzioni tra utilizzi diversi dell’analisi della performance. Una distinzione importante è quella tra *accountability* verso il basso e verso l’alto. Sia JTPA sia GPRA sono strumenti di rendicontazione verso l’alto, cioè verso livelli di governo sovraordinato: la misurazione viene effettuata in quanto imposta da un livello di governo superiore a cui l’organizzazione deve rispondere della propria performance (secondo profili stabiliti dall’autorità superiore o concordati dalle due parti). In altre situazioni, indicatori di performance possono essere utilizzati per rendere conto verso il basso, cioè verso gli utenti di un servizio, siano essi cittadini, imprese, o livelli di governo inferiori o laterali.

Un esempio di strumento di *accountability* verso il basso è fornito dalle Carte dei Servizi: al loro interno vengono utilizzati indicatori di performance e vengono fatti confronti tra performance osservate e standard preordinati. A questo proposito Castegnaro, De Ambrogio e Lo Schiavo (1999) discutono di un’interessante distinzione, quella tra standard specifici e standard generali. I primi sono verificabili dal singolo utente, che può confrontare le caratteristiche qualitative della singola prestazione ricevuta con un valore puntuale (“il tempo di risposta ad un’istanza non deve superare il 5 giorni lavorativi”) e decidere se agire (protestare, chiedere un rimborso). Nel caso degli standard generali (“la proporzione di utenti che ricevono risposta oltre i 5 giorni lavorativi non deve superare il 10%”) non può essere il singolo ad agire per il non rispetto dello standard, ma saranno istanze organizzate a poterlo fare.

Quest’ultimo esempio si ricollega ad un punto più generale, che va sempre tenuto presente: l’informazione sulla performance va generata in modo da essere correttamente interpretabile dall’utilizzatore rilevante. Un indicatore di performance può fornire informazione utile per l’utilizzatore A in un certo contesto, ma può essere irrilevante per l’utilizzatore B nello stesso contesto. Per questo è particolarmente importante riflettere con cura su chi utilizzerà quali indicatori prima di cimentarsi a definirli ed implementarli. La mancanza di questa riflessione porta con grande probabilità a generare quantità di numeri inutili.

Un esempio relativo alla valutazione dell’Università può servire ad illustrare questo punto. Tra i molti aspetti della performance di un’Università che possono essere valutati ci limitiamo qui a considerare la valutazione della didattica dei docenti, recentemente sancita per legge anche in Italia. Una tipologia di indicatori di performance della didattica universitaria può essere ricavata mediante questionari rivolti agli studenti, in cui li si invita ad esprimere giudizi su vari aspetti della didattica impartita da ogni singolo docente in ogni singolo corso. Indicatori di performance del tipo “percentuale di studenti che giudica come buona/ottima la chiarezza espositiva del docente” possono essere ricavati da questa fonte. La questione che va subito chiarita è: chi può proficuamente utilizzare gli indicatori che possono essere ricavati dallo spoglio dei questionari? Per riprendere la domanda posta all’inizio del capitolo, quali soggetti potranno reagire alle informazioni prodotte da questi indicatori e in quale modo potrebbero venire indotti a modificare i loro comportamenti/decisioni? In questo pur semplice contesto, la risposta alla domanda non è banale. A nostro parere l’utilizzo primario di questi indicatori è nell’*eseguire confronti* tra i docenti della stessa Facoltà o

<sup>8</sup> “*Usare la misurazione della performance solo a scopo di accountability verso livelli più alti di governo sarebbe un enorme spreco. I dati sulla performance devono essere usati per migliorare la gestione dei programmi di intervento.*” (Hatry, 1999).

corso di laurea. Da questi confronti emergerà un giudizio sulla performance di ciascun docente relativa a quella degli altri. Tale informazione ha un valore esclusivamente “locale” ed un utilizzo locale: è il singolo docente o il gruppo di docenti “valutati” che possono utilizzare proficuamente gli indicatori e trarne le debite conseguenze per il modo con cui impostano la didattica.

Si possono tuttavia realizzare aggregazioni di questi indicatori a livelli più elevati, per condurre confronti tra Facoltà o tra Atenei, o addirittura per ottenere un valore aggregato dell’indicatore a livello nazionale. Tuttavia ci chiediamo: ha senso la produzione di questi indicatori aggregati? In questo caso specifico, l’aggregazione è a nostro parere di dubbia utilità: il giudizio espresso da ciascuno studente su un docente è sempre relativo agli altri docenti che lo studente è in grado di osservare, cioè appartenenti alla stessa Facoltà o corso di laurea. Quindi i confronti tra i giudizi espressi in media dagli studenti dell’Ateneo A con quelli espressi in media dagli studenti dell’Ateneo B sono di scarsa interpretabilità: manca un punto di riferimento comune affinché giudizi espressi da studenti di Atenei diversi siano davvero confrontabili. Gli studenti dell’Ateneo A potrebbero essere mediamente più esigenti nelle loro aspettative e quindi dare giudizi più severi, a cui non corrisponderebbe però necessariamente una performance obiettivamente più scadente da parte dei docenti dell’Ateneo A. Quindi avrebbe poco senso individuare come utilizzatori di questi indicatori di valutazione della performance i Rettori delle Università o addirittura il Ministero dell’Università: fornire a questi soggetti un’informazione aggregata ma difficilmente interpretabile sarebbe un pessimo servizio. In questo caso, l’utilizzatore degli indicatori va individuato a livello locale, dove l’informazione prodotta può essere correttamente interpretata.

Un altro esempio può essere fatto relativamente alla formazione professionale. La “percentuale di diplomati che trova lavoro entro un anno dalla formazione” è senza dubbio un utile indicatore di performance di un corso, di un centro, o dell’intero sistema regionale o nazionale di formazione. Tuttavia, utilizzatori diversi richiedono un modo diverso di strutturare questa informazione. Un potenziale studente che debba decidere a quale centro di formazione iscriversi ha interesse a poter confrontare questo centro con gli altri, ma questo confronto non deve essere depurato dalle differenze nelle condizioni di partenza degli allievi: allo studente interessa frequentare la scuola frequentata dagli studenti più occupabili. Viceversa, l’Assessorato regionale alla formazione dovrebbe essere interessato a confrontare la “percentuale di diplomati che trova lavoro entro un anno dalla formazione” di ciascun centro con uno standard che tenga conto delle condizioni “oggettive” in cui opera il centro e che sfuggono al suo controllo, come nel caso di JTPA. Questo perché l’Assessorato vuole giudicare la performance attribuibile alle azioni e decisioni del singolo centro, depurandola dei condizionamenti esterni, mentre il singolo studente ha interesse anche ad essere informato sui condizionamenti esterni che influiscono sul risultato finale prodotto dal singolo centro.

Concludendo su questo punto, per avere successo un sistema di misurazione della performance deve partire da premesse chiare sulle sue finalità e sui suoi destinatari: soprattutto deve stabilire un legame credibile tra l’informazione prodotta e l’utilizzo che i destinatari possono farne. Restare agnostici su questi punti, delegando a dopo la misurazione la definizione dei destinatari, è una condanna quasi sicura all’inutilità della successiva misurazione.

### **3.2. La costruzione degli indicatori: cosa si vuole e può misurare?**

Il processo di costruzione di un sistema di indicatori di performance dovrebbe essere letto come un percorso ad ostacoli; ogni ostacolo è rappresentato da una domanda a cui il valutatore deve *necessariamente* dare una risposta per passare all’ostacolo successivo. Il premio per chi termina con successo questo percorso è l’individuazione di indicatori di performance validi, credibili e, quindi, potenzialmente utili al processo decisionale. La prima fondamentale domanda da porsi prima di avviare qualsiasi attività di misurazione è la seguente: *come definisco, e quindi su quale base giuridico, il successo dell’organizzazione?*

Sebbene possa apparire un quesito banale, la risposta a questa domanda non può mai essere data per scontata. Uno degli elementi che accomunano le tre esperienze che abbiamo esaminato è proprio la grande attenzione con la quale sono state definite le dimensioni di successo sulle quali basare il giudizio sulla *performance* dell’organizzazione. Ovviamente i percorsi seguiti nelle tre esperienze sono tra loro molto diversi e dipendono essenzialmente dalla natura delle stesse organizzazioni e dai differenti contesti all’interno dei

quali esse si muovono. Nell'esperienza di GPRA le Agenzie federali giungono all'identificazione dei risultati delle proprie attività attraverso un elaborato processo di *strategic planning* che coinvolge addirittura gli *stakeholders* dell'organizzazione ed impiega alcuni anni prima di giungere a conclusione; in JTPA, più semplicemente si identifica, dopo alcuni tentativi, la principale dimensione di successo con la probabilità di trovare lavoro da parte dei formati (all'avvio del programma veniva data grande importanza anche al fattore costo); Hatry affronta il problema sottolineando l'esigenza per i *manager* pubblici di riconoscere gli *outcome* da tenere sotto controllo (“*What outcome should be tracked?*”).

Qualunque sia il modo con cui si arrivi a tale definizione, la concreta possibilità di giudicare il successo di un'organizzazione passa sempre dalla chiara individuazione delle finalità che essa persegue, intese come “*situazioni desiderabili verso le quali l'organizzazione dovrebbe essere diretta*” (riduzione del rischio di perdite di vite umane nel corso di un incendio; aumento dell'occupabilità dei disoccupati di lungo periodo; diminuzione dei costi necessari a garantire un determinato servizio.) Il maggiore o minore successo dell'organizzazione dipende quindi da quanto essa è vicina, o si sta avvicinando, a tali situazioni; e qui nasce la seconda domanda al quale il valutatore deve dare risposta: *come posso misurare se l'organizzazione si sta muovendo nella direzione desiderata?*

Nasce cioè il problema di tradurre delle ipotetiche dimensioni di successo in quantità misurabili e, quindi, in indicatori. Ad esempio, nel caso della Protezione Civile il successo dell'organizzazione nel ridurre il rischio di perdite di vite umane durante gli incendi viene direttamente misurato dal numero di morti avvenute in simili circostanze, in un certo periodo di tempo; in JTPA la capacità di ogni agenzia per l'impiego di migliorare le condizioni occupazionali viene misurata dalla percentuale di occupati tra gli allievi, a tredici settimane dalla conclusione del corso, e dalla retribuzione settimanale di coloro che sono occupati. La traduzione delle dimensioni di successo in “*misure*” però non è mai un'operazione univoca e presuppone sempre una scelta da parte del valutatore; la distanza dell'organizzazione da situazioni ritenute desiderabili può essere misurata facendo infatti riferimento a diverse entità. Nel primo caso portato ad esempio, invece di calcolare il numero di morti avvenute durante gli incendi, si potrebbe considerare il numero di persone salvate dagli incendi grazie alle operazioni di evacuazione e di soccorso predisposte dalla Protezione Civile; allo stesso modo, nel secondo caso, potrebbe essere impiegata come misura di performance la quantità di ore giornaliere lavorate in media dai formati, nell'anno successivo alla conclusione del corso di formazione. In ogni caso la scelta di adottare un determinato indicatore per misurare un aspetto particolare della performance dovrebbe essere sempre ben argomentata e, comunque, posta a successiva verifica; l'adozione dell'indicatore sbagliato, perché non valido<sup>9</sup>, potrebbe distorcere drammaticamente il giudizio sulla performance dell'organizzazione.

Ma i problemi legati alla misurazione non terminano con la scelta dell'indicatore più valido; una volta effettuata tale scelta, occorre porsi la terza domanda che mette in luce un ulteriore problema legato alla raccolta delle informazioni: *in che modo posso generare le informazioni che mi servono?*

Di solito è possibile fare riferimento a tre diverse fonti informative: a) i dati generati nel corso delle abituali attività svolte dall'organizzazione; b) i dati raccolti mediante indagini *ad hoc* sugli utenti dei servizi dell'organizzazione o sugli operatori stessi; c) i dati provenienti da sistemi informativi preesistenti e non legati all'attività svolta dall'organizzazione. Naturalmente, dato il carattere limitato di questo articolo, non è possibile esaminare le notevoli differenze esistenti tra queste tre fonti informative in quanto ad accuratezza, completezza, comparabilità<sup>10</sup>, costo e cadenza temporale dei dati prodotti; tantomeno è possibile soffermarsi sulle cautele che è necessario seguire nella raccolta di tali informazioni, in quanto esse sono caratteristiche e peculiari allo strumento di rilevazione utilizzato. Ad esempio la rilevazione di informazioni attraverso un

<sup>9</sup> Si tratta del problema, tipico di ogni scienza sociale, della validità delle misure scelte per rappresentare un fenomeno. Prima di adottare un qualsiasi strumento di misurazione è necessario porsi alcune domande: “*che cosa misura? I dati che fornisce sono rilevanti per le caratteristiche a cui si è interessati? Sino a che punti le differenze (riscontrate) rappresentano differenze reali delle caratteristiche che stiamo cercando di misurare? Sino a che punto riflettono anche l'influenza di altri fattori?*” (Selltiz e altri, 1976).

<sup>10</sup> “*I dati utilizzati in una valutazione dovrebbero possedere tre caratteristiche: (1) essi dovrebbero essere ragionevolmente accurati; (2) essi dovrebbero essere ragionevolmente completi; e (3) poiché la valutazione si basa perlopiù su confronti tra dati raccolti in periodi di tempo differenti, e possibilmente per diversi segmenti della popolazione, essi dovrebbero essere comparabili*” (Hatry ed altri, 1981).

questionario, compilato da un intervistatore pone problemi completamente diversi dal trattamento di dati amministrativi, raccolti in precedenza dall'organizzazione per altri scopi. Qualunque sia lo strumento utilizzato vale una raccomandazione generale: dubitare sempre della veridicità del dato e cercare comunque una sua convalida attraverso indagini suppletive.

Prima di adottare la misura scelta come indicatore di performance, occorre ancora superare l'ultimo ostacolo: *quali comportamenti (indesiderati) può innescare l'utilizzo dell'indicatore?*

L'organizzazione può sempre rispondere in modo imprevisto e tendenzialmente indesiderato all'impiego di uno o più indicatori concepiti per giudicarne la performance. In linea generale l'organizzazione, che si sente controllata, può dare due tipi di risposte negative:

- una volta che abbia compreso in che modo è stato costruito l'indicatore e il meccanismo che lo regola, l'organizzazione può tentare di migliorarne il valore, manipolando esclusivamente le informazioni raccolte, senza mutare di fatto la propria performance; oppure
- può adottare comportamenti tattici che hanno l'effetto di migliorare il valore assunto dall'indicatore a fronte, però, di peggioramenti della propria performance.

Una distorsione del primo tipo è quasi sempre possibile quando la raccolta delle informazioni è effettuata dagli stessi operatori che lavorano all'interno dell'organizzazione (di fatto "i giudicati") e la definizione del tipo di informazione raccolta non è molto precisa e lascia spazio alla libera interpretazione del rilevatore. Un simile caso non è così infrequente come si potrebbe pensare. Ipotizziamo ad esempio di valutare la performance delle agenzie per l'impiego attraverso l'utilizzo di un indicatore costruito come un rapporto tra il numero di utenti "contattati" o "raggiunti" dall'agenzia e il numero di "utenti potenziali" (ad esempio il numero di disoccupati iscritti alla lista di collocamento)<sup>11</sup>; il limite di tale indicatore risiede proprio nell'interpretazione che può essere attribuita al termine "contattato". Se vengono lasciati ampi margini di discrezionalità, all'operatore che rileva il dato, nella scelta del significato da assegnare al termine "contattato", è chiaro che egli lo interpreterà nel modo a lui più favorevole, facendo aumentare artificialmente il valore dell'indicatore al fine di dimostrare una buona performance.

La distorsione del secondo tipo è ancora più grave in quanto non solo mina l'attendibilità delle informazioni raccolte, ma finisce per causare addirittura un peggioramento nella performance stessa dell'organizzazione: l'esatto contrario di quanto si intende ottenere con l'impiego di indicatori. Un buon esempio di questo tipo di distorsione è dato dall'esperienza di JTPA. Inizialmente il sistema di misurazione di performance delle agenzie prevedeva l'adozione di due indicatori: il costo medio per allievo e, come abbiamo visto, la percentuale di occupati tra i formati dopo un certo periodo di tempo. Dopo alcuni anni di operatività del sistema ci si è accorti di due fatti spiacevoli: innanzitutto l'adozione del costo medio per allievo, allo scopo di tenere sotto controllo i costi del programma, induceva le agenzie ad accorciare la durata dei corsi e a ridurre l'intensità per aumentare il numero degli allievi a parità di costo totale: ciò naturalmente diminuiva il valore dell'indicatore a scapito però della qualità della formazione impartita ai singoli studenti. Inoltre, al fine di mantenere elevata la percentuale di occupati tra i formati, alcune agenzie cercavano di scoraggiare l'iscrizione degli individui più svantaggiati, che li avrebbero penalizzati in sede di misurazione, in quanto essi avevano minore probabilità di trovare un'occupazione al termine del corso (scrematura). Il verificarsi di tali comportamenti ha condotto perciò all'eliminazione del primo indicatore dal sistema di misurazione della performance e alla correzione nelle modalità di utilizzo del secondo indicatore<sup>12</sup> (come abbiamo già descritto nella sezione 2.2 attraverso un modello di regressione che tenga conto delle caratteristiche degli utenti).

Riepilogando nella costruzione degli indicatori occorre: (i) identificare dimensioni rilevanti di successo; (ii) scegliere delle quantità misurabili che le rappresentino correttamente; (iii) individuare fonti informative attendibili; (iv) fare attenzione alle possibili risposte negative messe in atto dall'organizzazione.

---

<sup>11</sup> Nella letteratura sugli indicatori un rapporto di questo tipo viene solitamente denominato come *grado di copertura* o *di soddisfazione della domanda potenziale* (Farneti, 1996; Gori e altri, 1999).

<sup>12</sup> Come abbiamo già descritto nella sezione 2.2, lo *standard*, con il quale si confronta la *performance* della singola agenzia, viene corretto mediante il ricorso ad un modello statistico che tiene conto delle reali caratteristiche degli utenti.

### 3.3. Come si giudica la performance osservata?

Giudicare significa innanzitutto fare confronti tra qualcosa che osserviamo (un comportamento, una situazione, un prodotto) e qualcosa che dovrebbe o potrebbe essere. A seconda del termine di paragone prescelto, produciamo dei giudizi con una valenza e un'interpretazione diversa. Ma qualche termine di paragone è comunque sempre necessario: senza di esso non si ha un giudizio, ma una semplice rappresentazione di ciò che si osserva. Nell'ambito della misurazione della performance, la mancata individuazione di termini di paragone rischia di produrre numeri che restano inutilizzati o sottoutilizzati in quanto non interpretabili.

L'esempio di JTPA è a nostro parere illuminante: sono stati individuati soltanto due indicatori per giudicare la performance delle agenzie locali per l'impiego, mentre tutta l'attenzione di chi ha disegnato il sistema di valutazione si è spostata sulla procedura per fissare gli standard di performance, cioè i termini di paragone. E' vero che la particolare procedura utilizzata in JTPA non può essere generalizzata a tutte le situazioni. Ma quello che a noi interessa è sottolineare la sensibilità di fondo che caratterizza l'approccio: invece di impegnarsi nella spasmodica ricerca di tutti i possibili "descrittori" della performance, ci si è limitati a fornirne una descrizione molto parsimoniosa ma ci si è impegnati a cercare un modo per "far parlare" i pochi indicatori prescelti.

Le tre esperienze statunitensi mostrano come l'adozione di termini di confronto vari molto da caso a caso. Gli indicatori adottati nei *performance plans* previsti da GPRA vengono confrontati soprattutto con valori passati della performance e con target prestabiliti. Gli standard di JTPA implicano un confronto con un termine di paragone che deriva dalla fissazione di un target, corretto per tenere conto di ciò che è lecito aspettarsi in base alle condizioni in cui opera un particolare organizzazione. Il manuale di *Performance Measurement* indica invece un ampio spettro di possibili termini di paragone: da semplici confronti interni all'organizzazione (tra unità operative, tra categorie di utenti) a confronti con termini di paragone esterni all'ambito spazio-temporale in cui sono misurati gli indicatori (performance passate, performance di altre organizzazioni, target numerici, standard condivisi.)

Questo ampio spettro di possibilità può lasciare disorientati. Come si scelgono dunque i termini di paragone per condurre i confronti con i valori osservati degli indicatori? Non è possibile dare una regola generale: dipende in modo cruciale da quale deve essere l'utilizzo della valutazione. E' possibile tuttavia delineare alcuni criteri per orientarsi ad affrontare il problema.

A questo scopo abbiamo individuato quattro possibili utilizzi dell'analisi di performance, che a loro volta si collegano a quattro tipi diversi di confronti. Fondamentalmente, l'analisi di performance può servire a:

- segnalare anomalie nel funzionamento di un'organizzazione (confronto con standard minimi);
- stimolare l'organizzazione al miglioramento, confrontandosi con livelli alti di performance raggiunti da altri (confronto con *best practices*);
- verificare il progresso compiuto nel risolvere i problemi (confronto con i livelli di partenza);
- verificare quanto si è realizzato (confronto con quanto è stato programmato).

Questi utilizzi non si escludono a vicenda: anzi, un'analisi di performance può ambire a rispondere a tutte e quattro le esigenze delineate. Tuttavia, la distinzione resta utile per motivare la scelta dei diversi termini di paragone.

#### (CONFRONTI PER) SEGNALARE ANOMALIE DI FUNZIONAMENTO

Uno dei motivi più comuni per condurre una misurazione della performance è l'idea che esista uno spettro di performance accettabili da parte di una certa organizzazione e che la misurazione serva sostanzialmente a segnalare anomalie, cioè prestazioni scadenti, comportamenti scorretti, deviazioni dalla norma. All'individuazione delle anomalie fa seguito un qualche tipo di azione correttiva, o almeno la possibilità di un'azione correttiva. L'assunto è che senza la segnalazione da parte del sistema di misurazione della performance, l'anomalia sopravviverebbe indisturbata e nessuna azione correttiva sarebbe intrapresa. In questo quadro, la scelta dei termini di paragone risponde alla logica di individuare, per ciascun indicatore, *un livello minimo accettabile* di performance: ad esempio, un costo massimo per prestazione, un tempo

massimo di attesa, un numero minimo di utenti per operatore, un numero massimo di allievi per docente. Al di sotto di questi minimi (o al di sopra dei massimi), la performance è da considerarsi anomala e un'azione correttiva si rende necessaria.

Due elementi sono necessari per arrivare alla determinazione di questi standard minimi (o massimi a seconda dei casi): innanzitutto la conoscenza della distribuzione delle performance in organizzazioni simili o nella stessa organizzazione per un periodo passato sufficientemente lungo; e poi un elemento ineliminabile di arbitrarietà nello stabilire cosa rappresenta un "livello accettabile". Occorre sottolineare che i due elementi vanno combinati e non si dovrebbe mai procedere con uno solo: gli standard minimi non possono essere scelti in modo *del tutto* arbitrario, con il rischio di segnalare come anomali troppi (o troppo pochi) casi; d'altro canto è difficile (al di fuori di processi produttivi di tipo industriale o comunque altamente standardizzato) stabilire in modo puramente statistico cosa rappresenti un comportamento anomalo. Il caso dei *performance standard* di JTPA è un esempio di individuazione di standard minimi, che combina un elemento di arbitrarietà con la conoscenza dell'intera distribuzione delle performance derivante dall'osservazione di tutte le organizzazioni in un certo anno.

*(CONFRONTI PER) STIMOLARE IL MIGLIORAMENTO*

Per quanto "stimolare il miglioramento" possa essere considerato l'obiettivo generale della misurazione della performance, ha anche senso considerarlo come un modo particolare di generare termini di paragone. Qui i valori degli indicatori non vengono confrontati con standard minimi, bensì con valori che rappresentano il meglio o il massimo che sia "ragionevole" raggiungere rispetto a quell'indicatore. Anche in questo caso per generare i termini di paragone occorre combinare un elemento di arbitrarietà con la conoscenza della distribuzione delle performance di altre organizzazioni. Qui però si guarda alla "coda alta" della distribuzione, invece che alla sua "coda bassa": si guarda ai casi di eccellenza, ai casi di *best practice*, individuati in modo inevitabilmente soggettivo.

Anche in questo caso quindi non si deve fare a meno di tener conto della reale distribuzione delle performance. Certo, sarebbe possibile anche fissare *target* del tutto arbitrari e in certe situazioni ciò può riuscire a stimolare la performance. Ma come regola generale è importante che il *target*, il livello da raggiungere, sia in qualche modo credibile, cioè sia stato realmente prodotto da qualche organizzazione che opera in condizioni non troppo dissimili da quella valutata. E' ragionevole ipotizzare che esista una correlazione positiva tra la credibilità di un *target* e la sua efficacia nello stimolare il miglioramento. Se un *target* non è fondato sull'esperienza realizzata da altri e quindi rischia di essere percepito come poco realistico, minore sarà la sua efficacia nello stimolare l'organizzazione verso il miglioramento.

*(CONFRONTI PER) VERIFICARE IL PROGRESSO COMPIUTO NEL RISOLVERE I PROBLEMI*

La motivazione dell'analisi della performance è diversa nel caso in cui intenda verificare il cambiamento nei fenomeni su cui l'azione pubblica vuole incidere: ad esempio, di quanto si è ridotto il tasso di criminalità, di quanto è aumentata la velocità di circolazione del traffico, di quanto è diminuito l'inquinamento atmosferico, di quanto è migliorato il livello di pulizia delle strade. In questi casi il termine di paragone tende ad essere ricavato dall'esperienza passata: si badi, non necessariamente dal periodo immediatamente precedente, ma dal momento in cui ci si è posti l'obiettivo di cambiamento. In inglese si parla di confronti con una *baseline*.

Ovviamente, questo tipo di confronti soffre di una forte debolezza interpretativa, che deve consigliare di usarli con molta cautela. Molteplici sono le cause che determinano l'evolversi di un fenomeno nel tempo e l'attività di un'organizzazione pubblica che tenta di modificare il fenomeno è solo una di queste possibili cause. Per questo il semplice paragone tra un punto di partenza e un punto di arrivo è raramente una valida stima dell'effetto netto dell'azione pubblica. Queste considerazioni non vogliono essere un deterrente all'utilizzo di questi termini di paragone, ma una cautela nel modo di interpretarli. L'esempio di GPRA in questo caso è utile: gran parte dei confronti condotti in quella sede sono confronti di un indicatore con una *baseline*, allo scopo di mostrare il progresso compiuto verso l'obiettivo di cambiamento. Però abbiamo visto come il GAO solleciti le Agenzie ad arricchire l'analisi con informazioni sui possibili fattori esterni che influenzano il valore dell'indicatore e possono impedire il raggiungimento dell'obbiettivo.

*(CONFRONTI PER) VERIFICARE IL REALIZZATO RISPETTO AL PROGRAMMATO*

Il confronto tra ciò che è stato realizzato e ciò che è stato programmato viene talvolta indicato quasi fosse un *sinonimo* di analisi di performance. In quest'ottica, obiettivo significa essenzialmente "cose da fare": gli indicatori di performance assumono allora la forma di operazioni espletate, casi esaminati, utenti serviti, progetti realizzati e i termini di paragone con cui i valori osservati degli indicatori vengono confrontati sono espressi in termini di operazioni programmate, casi previsti, utenti programmati, progetti presentati. La misurazione della performance si ridurrebbe, se si assecondate del tutto questa visione, ad una verifica di ciò che è stato fatto da un'organizzazione. Ci lascia molto perplessi l'idea di "onniscienza del programmatore" che sottende a questa visione, in particolare l'idea che tutto sia programmabile e quindi al valutatore non resti che un'opera di verifica puntuale della realizzazione del programma. Infatti si definisce talvolta "efficacia" come "rapporto tra realizzato e programmato". Quanto più la realizzazione si avvicina ai valori programmati, tanto più alta sarebbe l'efficacia del servizio.

Riteniamo che questa visione della valutazione sia molto parziale e riduttiva. Senza dubbio gli obiettivi e gli impegni fissati negli atti di programmazione possono essere fonte di termini di paragone da utilizzare in un'analisi di performance, ma non certo l'unica fonte. Il confronto tra realizzato e programmato può produrre informazione utile in talune circostanze, ma tale informazione va posta nella giusta luce: si tratta pur sempre di uno scostamento rispetto ad un livello programmato, il che non consente alcun giudizio sul valore, sull'utilità, quindi sulla reale efficacia dell'attività svolta.

#### **4. Riflessioni sull'uso degli indicatori in Italia**

Nella discussione precedente abbiamo fatto costantemente riferimento ad indicatori di performance, piuttosto che usare la definizione, più comune in Italia, di "indicatori di efficienza ed efficacia". Questa scelta non è casuale, ma trova motivazione nel disagio che chi scrive prova per l'uso inflazionato e poco meditato che di questi termini si fa oggi in Italia, in particolare all'interno del processo di costruzione di batterie di indicatori.

Il punto di partenza del nostro ragionamento è che tali batterie di indicatori soffrono di un difetto di nascita, dovuto alla scarsa attenzione prestata, al momento del loro concepimento, alla loro concreta utilizzabilità a fini decisionali. Quella che segue è ovviamente una semplificazione, applicabile forse solo ad alcune situazioni. Lasciamo al lettore giudicare quanto la nostra semplificazione si avvicini alle esperienze di cui sia stato testimone.

Il punto di partenza nella costruzione di batterie di indicatori pare spesso essere una lista di quantità osservabili che concernono l'attività dell'organizzazione o del servizio di cui si vuole giudicare la performance: ad esempio, utenti serviti, risorse impiegate, prestazioni offerte, tempi di risposta, domande presentate, domande accettate. Si tende cioè a partire dai dati disponibili o da quelli che possono facilmente essere raccolti nell'ambito dell'attività dell'organizzazione, invece che interrogarsi realmente su quali sono i requisiti informativi rilevanti per il processo decisionale.

Il passo immediatamente successivo consiste nella costruzione degli indicatori, che non sono altro che rapporti tra coppie di quantità. Spesso tali rapporti rappresentano semplici "normalizzazioni", ad esempio nel caso "prestazioni/abitanti", per cui un valore alto o basso dell'indicatore non consente di per sé alcun giudizio sul servizio. In altri casi il valore del rapporto implica invece un giudizio, ma non sempre di segno univoco: ad esempio, i valori del rapporto "domande accettate/domande presentate" a seconda del contesto possono rappresentare un giudizio positivo o un giudizio negativo (se si sta giudicando la capacità di selezionare, un valore basso può essere positivo, se si sta giudicando la qualità delle domande presentate, un valore alto diventa positivo.)

Il passo finale consiste nell'assegnare a ciascun indicatore così costruito un'*etichetta* corrispondente ad una caratteristica desiderabile dell'azione pubblica. Tali caratteristiche sono solitamente astratte e indefinite: tra esse primeggiano le "tre E" (efficienza, efficacia, economicità), ma si trovano anche rilevanza, pertinenza, impatto, produttività e molte altre. Il caso dell'efficacia è emblematico in quanto a confusione terminologica: questo termine assume una gran quantità di significati diversi, che vanno da semplice sinonimo di "qualità"

ad accezioni più impegnative come “grado di soddisfacimento della domanda” nel caso di servizi potenzialmente sottoposti a razionamento, a “capacità di realizzare quanto previsto”, al più stringente “capacità di determinare gli effetti desiderati”.

L'operazione di *etichettatura degli indicatori* è fondamentale all'interno di questo approccio, anzi a nostro parere ne rappresenta il passo essenziale. L'indeterminatezza di significato di molte etichette non è un deterrente al loro uso. Quando un indicatore è stato etichettato assume una dignità che altrimenti non avrebbe. Un rapporto del tipo “utenti serviti/utenti potenziali” assume tutta un'altra dignità nel momento in cui viene etichettato come “indicatore di efficacia”. Non solo una più alta dignità, ma anche una più immediata interpretabilità: essendo l'efficacia una cosa sempre desiderabile, valori alti del rapporto sono sempre desiderabili, e la massima efficacia si raggiunge quando il valore raggiunge il 100 per cento. Non è detto che servire una maggior percentuale di utenti potenziali sia un indice di maggior capacità di soddisfare la missione che il servizio si prefigge: ma nel momento in cui il rapporto utenti serviti/utenti potenziali viene etichettato come “di efficacia”, il giudizio positivo diventa immediato. Le stesse considerazioni si possono fare ad esempio per rapporti del tipo “atti realizzati/atti previsti”: una volta che sia etichettato come “indicatore di efficacia”, l'ottimo si raggiunge solo puntando al 100 per cento. Trattandosi di efficacia, il giudizio è automatico e immediato.

Con questo terzo passo, cioè l'etichettatura degli indicatori, si conclude spesso la parte di disegno concettuale del sistema di misurazione/valutazione. Non resta che passare alla fase operativa, trovare i numeri per “riempire” gli indicatori, e la valutazione dell'organizzazione discenderà dalla semplice osservazione dei valori così ottenuti. La loro interpretabilità è scontata e non problematica, perché il significato ai numeri così ottenuti deriva dalle etichette che sono state loro attribuite.

Quella appena presentata può sembrare una caricatura “provocatoria” della costruzione di batterie di indicatori. In una certa misura lo è, ma come tutte le caricature crediamo contenga un elemento di verità. Le operazioni essenziali sono: l'identificazione di ciò che è misurabile, la definizione di tutti i possibili rapporti tra quantità misurabili, la loro etichettatura mediante categorie astratte. A nostro parere mancano due passaggi, essenziali affinché tutto ciò abbia significato. A monte mancano l'aggancio con il processo decisionale; una chiara visione ed esplicitazione del perché si vuole misurare la performance; una definizione attenta della missione del servizio e delle sue dimensioni di successo. A valle della costruzione degli indicatori manca spesso uno sforzo serio per individuare dei termini di paragone che diano significato alle misurazioni effettuate e le rendano interpretabili a fini valutativi.

Quella che speriamo di aver offerto in questo lavoro, partendo dall'esame di alcune esperienze statunitensi, è una visione diversa dell'analisi di performance, in cui la costruzione degli indicatori è solo un singolo momento di un processo molto più complesso ed articolato. Ci auguriamo che questa nostra “provocazione” possa stimolare la riflessione e il dibattito.

### **Bibliografia**

- Barnow, B., “The Effect of Performance Standards on State and Local Programs” in Manski e Garfinkel, (a cura di) *Evaluating Welfare and Training Programs*, Cambridge, MA: Harvard University Press, 1992.
- Barnow, B., “Exploring the Relationship Between Performance Management and Program Impact: A Case Study of the Job Training Partnership Act” in *Journal of Policy Analysis and Management*, vol. 19, no. 1, 2000.
- Castegnaro C., De Ambrogio U., e Lo Schiavo M., “La costruzione degli standard di qualità dei servizi sociali. Una proposta di lavoro”, in *Prospettive Sociali e Sanitarie*, n. 21-22, 1999.
- Dejean, J., Bourassin, V., e Mortreux, K., “Culture de l'Evaluation et Fascination pour les Indicateurs” in *Politiques et Management Public*, vol. 16, no. 2, 1998.

- Dickinson, K., *Evaluation of the Effects of JTPA Performance Standards on Clients, Services and Costs*, Washington, DC: National Commission for Employment Policy, 1988.
- Farneti, G., Mazzara, L., Savioli, G., *Il sistema degli indicatori negli enti locali*, Torino, Giappichelli, 1996.
- Federal Emergency Management Agency, *Fiscal Year 1999 Annual Performance Plan*, Washington D.C., 1998.
- General Accounting Office, *Executive Guide: Effectively Implementing the Government Performance and Result Act*, Washington D.C., 1996.
- General Accounting Office, *Results Act: Observations on the Federal Emergency Management Agency's Draft Strategic Plan*, Washington D.C., 1997.
- General Accounting Office, *Results Act: Observations on the Federal Emergency Management Agency's Fiscal Year 1999 Annual Performance Plan*, Washington D.C., 1998.
- General Accounting Office, *Results Act: Observations on the Federal Emergency Management Agency's Fiscal Year 1999 Annual Performance Plan*, Washington D.C., 1999.
- Gori, E., Vittadini G., *Qualità e valutazione nei servizi di pubblica utilità*, Milano, Etas, 1999.
- Hatry, H., *Performance Measurement: Getting Results*, Washington, DC: The Urban Institute Press, 1999.
- Hatry H. e altri, *Practical Program Evaluation for State and Local Governments*, Washington D.C., The Urban Institute, 1981.
- Newcomer, K., "Using performance measurement to improve programs", in *New Directions for Evaluation*, N.75, Fall, Jossey-Bass Publishers, 1997.
- Perrin, B., "Effective Use and Misuse of Performance Measurement", in *American Journal of Evaluation*, vol. 19, no. 3, 1998.
- Selltiz C., Wrightsman L.J., Cook S.W., *Research Methods in Social Relations*, New York, Holt, Rinehart and Winston, 1976.