

I PRINCIPI DELLA VALUTAZIONE DELLA RICERCA

di Fiorella Kostoris Padoa Schioppa

L'articolo propone una tassonomia dei tipi di valutazione della ricerca e ricorda che quella ex post non è mai stata realizzata in Italia prima che vi si cimentasse il CIVR con riferimento al triennio 2001-2003. Sulla base in particolare di quell'esperienza, l'articolo cerca di rispondere a 5 domande: perchè impegnarsi in una siffatta forma di valutazione? È possibile farla bene? Se sì, con quali strumenti? Se sì, con quali formule o caratteristiche oggettive? Se sì, con quali tipi di valutatori o caratteristiche soggettive?

Mi riferirò ai principi generali e non ai dettagli, e parlerò di questioni attinenti la valutazione della ricerca su larga scala, non quella su un singolo prodotto sottoposto al giudizio, se ad esempio inviato per la pubblicazione ad una rivista con referaggio, o su un singolo ricercatore soggetto a concorso universitario.

Dopo una breve tassonomia sui tipi di valutazione della ricerca, vorrei dividere il mio intervento in 5 punti, chiedendo:

perché farla? È possibile farla bene? Se sì, con quali strumenti? Se sì, con quali formule o caratteristiche oggettive? Se sì, con quali tipi di valutatori o caratteristiche soggettive?

Tassonomia

Nel presente contesto, che chiarirò meglio nel seguito, un aspetto tassonomico è particolarmente rilevante.

Esso attiene al momento della valutazione rispetto a quello della ricerca: e allora si distingue la valutazione *ex ante*, *in itinere*, *ex post*. In Italia, fino ad un recente passato, si usava solo la prima, se e quando una qualche valutazione si faceva prima del finanziamento di un progetto, normalmente mai per un periodo superiore al triennio.

Tuttavia, è chiaro che, per progetti significativi, la durata della ricerca supera quella dell'iniziale allocazione delle risorse e nuovi finanziamenti si rendono necessari. Questo richiederebbe una valutazione *in itinere*, che per le nostre Università non si compie mai, ma che invece è la norma all'estero (ad esempio, negli USA); solo formalmente, prima di riaccedere alla stessa fonte di finanziamento, i ricercatori universitari spesso sono tenuti (e non sempre) a presentare un rapporto sulla ricerca compiuta, che normalmente però finisce nel dimenticatoio, se non nel più vicino cestino. La situazione è diversa

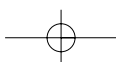
negli Enti di Ricerca, dove quasi sempre esiste un Consiglio Scientifico che ha il compito di monitorare il lavoro in corso ma, di solito, da un lato, lo fa in modo "user-friendly" e, dall'altro, non condiziona né il livello complessivo del finanziamento, né l'allocazione per specifiche aree. Solo recentemente - a partire dal Decreto MIUR del dicembre 2003 - è stato creato in Italia, con il CIVR, un meccanismo di valutazione *ex post* della ricerca per il periodo 2001-2003, che è iniziato con l'emanazione di linee guida decise da un *board* di 7 membri (cui ho l'onore di appartenere) e con l'autoselezione dei prodotti migliori da parte di tutte le Strutture di ricerca. L'esercizio dovrebbe concludersi a metà del 2006, avendo impegnato nel corso del 2005 20 *panel* (1 per ognuna delle 14 aree del CUN + 6 interdisciplinari), 150 panelisti, 6.000 esperti, circa 18.000 prodotti, 102 Strutture cui afferiscono 64.000 ricercatori, con un costo complessivo di 3 milioni e mezzo di euro.

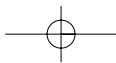
L'Onorevole Walter Tocci, nella sua ampia Relazione circa "I primi passi del Governo su Ricerca e Università" del 12 giugno 2006 scrive che "il metodo CIVR si è rivelato buono e con alcune messe a punto può essere esteso da subito agli anni successivi, al triennio 2004-2006 e poi dal 2007 per ogni annualità". Il neo-Ministro Mussi ha già chiesto al *board* del CIVR di continuare il lavoro di valutazione sul periodo 2004-2006, il che lascia bene sperare che l'esperimento sul 2001-2003 sia solo il primo di un sistema di *assessment* permanente, simile a quello inglese.

Nel seguito, tratterò esclusivamente dei principi della valutazione *ex post* della ricerca.

Perché farla?

Un motivo per valutare è già venuto fuori esplicitamente: è necessario monitorare il campo per individuare punti di eccellenza o carenze,





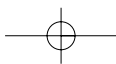
nei livelli e nelle dinamiche, e selezionare il meglio non tanto per conoscere, ma per indirizzare le risorse scarse del *budget* dedicato alla ricerca (che sappiamo tutti dover crescere dall'attuale 1,16% del PIL al 3% previsto da Lisbona). Questo ha detto, ad esempio, la ex Ministra Moratti quando ha indicato come avrebbe usato i risultati del CIVR per allocare 1/3 dei finanziamenti di ricerca destinati all'Università. E il programma dell'Unione – all'epoca all'opposizione – dopo aver affermato, in linea con la CRUI nel 2005, che servono “*la promozione della qualità in tutti gli Atenei tramite una valutazione continua ed efficace e lo sviluppo dell'eccellenza nelle migliori esperienze in atto nel sistema Università – Enti di Ricerca*”, prosegue specificando che la valutazione della ricerca ha “*funzioni di ripartizione di incentivi finanziari premiali dei punti di qualità del sistema*”, ma deve allo stesso tempo tener conto di “*obiettivi di riequilibrio territoriale*” e curare “*l'inserimento di opportuni indicatori di genere nel benchmarking delle istituzioni scientifiche*”. Si direbbe perciò che il programma effettivo del Governo attuale potrebbe non seguire esattamente le promesse “*business-like*” dell'ex-Ministra, perché, accanto all'obiettivo di incentivare l'efficienza, potrebbe comparire quello di compensare a fini equitativi le debolezze di alcune regioni, quali il Mezzogiorno, o del comparto femminile nel suo complesso.

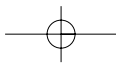
È possibile farla bene?

In Italia c'è un diffuso scetticismo sulla possibilità di selezionare un *target* prestabilito senza incorrere in distorsioni gravi e anche per questo c'è un'estesa preferenza nel *Welfare State* verso strumenti generalisti, normalmente indifferenti alle condizioni di bisogno o di merito e indipendenti da qualunque *means-testing*: questo è vero, che si tratti di proteggere la maternità dove non c'è il diritto (come nella stragrande maggioranza dei Paesi), ma il dovere di assentarsi per vari mesi, al fine di non stigmatizzare le donne più cagionevoli rispetto alle più sane, o che si tratti di offrire i prodotti farmaceutici, dove quelli di fascia A sono gratuiti per tutti, o l'istruzione superiore, dove le tasse universitarie sono solo debolmente correlate al reddito familiare (ma il numero di laureati lo è fortemente). C'è un largo malessere nei confronti della selezione darwiniana (o della concorrenza

che ne è l'equivalente nel mercato) e, si preferisce invocare protezioni per i potenziali perdenti piuttosto che rafforzare i socialmente deboli ma meritevoli, dando loro migliori opportunità di successo.

Al di là dei motivi ideologici per diffidare delle ragioni e dei metodi per distinguere i prodotti e le Strutture di ricerca di più elevato livello, appare chiaro che lo scetticismo sulla possibilità di individuarli senza distorsioni è però ormai ingiustificato, alla luce di tre risultati già emersi nella valutazione del CIVR: l'alta correlazione fra i giudizi indipendenti dei *referee reports*, esercitati su ogni prodotto di ricerca; se essi fossero stati arbitrari, tale correlazione sarebbe stata nulla o quasi. I giudizi che potevano essere dati erano di 4 tipi: eccellente, buono, accettabile, limitato. La *peer review* era esercitata da almeno due persone, in modo non coordinato. Ebbene, la concordanza completa (consenso unanime) dei giudizi riguarda, secondo un primo studio realizzato da Emanuela Reale ed altri due ricercatori di Chieti (“*New Frontiers in Evaluation*”, Vienna, 24-25 aprile 2006), una percentuale variabile fra il 43 e il 48% dei casi a seconda della disciplina, mentre la discordanza completa (ad esempio, un giudizio è eccellente e l'altro limitato o accettabile), con la conseguente necessità di rivolgersi ad un terzo *referee*, tocca una percentuale variabile fra il 9 e il 13% dei casi; solo molto raramente è concretamente avvenuto che qualche panelista del CIVR, purtroppo anche nel settore dell'economia, abbia dimenticato di verificare, come avrebbe dovuto, che sussistesse una fattuale separatezza fra i due *referee reports* e il *board* dei 7 membri è dovuto allora intervenire dall'alto per correggere questo *bias*, spedendo i prodotti ad altri esperti; la frequenza dell'unanimità nel consenso finale dei panelisti (che intervenivano a valle dei *referee reports*) è notevole: va da un minimo dell'86% dei casi nella disciplina economia, a valori del 94% nelle scienze naturali, al 97% in quelle umanistiche; infine, limitatamente ai prodotti costituiti da articoli pubblicati su riviste con impatto, è forte la correlazione fra il giudizio finale da parte dei *panel* sugli articoli sottoposti a valutazione e l'*impact factor* (IF) della rivista in cui essi sono pubblicati. Rimane la domanda se il giudizio della *peer review* realizzata dal CIVR sia di fatto indipendente dall'IF, e dunque se esso confermi la qualità degli indici bibliometrici, oppure ne sia condizionato. Il Presidente del panel di





economia, Franco Peracchi, in un suo recente "Giudizio sul CIVR" uscito il 15 maggio 2006 sulla voce.info, sostiene che in tale disciplina questo secondo caso non si manifesti.

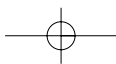
Fuori dall'esperienza del CIVR, l'indipendenza di giudizio delle almeno due persone chiamate alla *peer review* implica come condizione necessaria (ma non sufficiente) che le valutazioni dei *peers* vengano date per iscritto prima che essi si incontrino o magari incontrino i candidati. Così si giocano la reputazione e sono motivati a valutare con attenzione e onestà. Questa è la prassi, ad esempio, nell'NIH americano, con due giudizi per ogni domanda di *grant* di ricerca, o nel sistema francese dell'*agrégation*, utilizzato come una delle filiere per nominare professori ordinari, ma questa non è la prassi nei concorsi universitari italiani, dove di solito si scrive il "medaglione" dei candidati dopo l'incontro fra gli accademici giudicanti e spesso addirittura dopo il loro esame dei candidati.

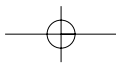
Se sì, con quali strumenti?

Il dibattito riguarda l'uso degli indici bibliometrici *versus* la *peer review*. Quest'ultimo è un *first best*, e non a caso è l'unico strumento utilizzato per la scelta della *Faculty* nelle migliori Università del mondo o per l'allocazione delle risorse nella NSF o nell'NIH americani, ma può venir distorta principalmente per ragioni di potere, e conseguentemente la discrezionalità rischia di diventare il peggiore degli strumenti. Perciò un *second best* è dato dagli indici bibliometrici, laddove è possibile applicarli (non in tutti i campi, per esempio non nelle materie umanistiche, e non per tutti i prodotti, per esempio non nei libri o nelle riviste senza IF). Naturalmente, un sistema misto è forse preferibile al *second best* ed esso, in qualche misura, è preso in considerazione dal *board* del CIVR per il futuro (al fine di poter valutare la totalità dei lavori di ricerca senza troppi costi e tempi troppo lunghi, anziché, come ora, limitandosi ad un numero di prodotti pari solo al 50% dei cosiddetti ricercatori equivalenti).

Alcuni indicatori bibliometrici sono, in proposito, migliori di altri. La CRUI, in un lavoro del 2005 ("Un aggiornamento sull'impatto della ricerca scientifica e tecnologica italiana in ambito internazionale, 1981-2004. Analisi preliminare" di Elena Breno et al.), suggerisce molto fortemente

che il *quotation index* sia più utile dell'*impact factor*, perché il numero di citazioni che un articolo ha avuto identifica meglio la sua intrinseca qualità, mentre l'*impact factor* è meno preciso, in quanto concerne la notorietà della rivista di pubblicazione nel suo insieme, costituendo una sorta di media dei *citation impact* dei diversi articoli. Tuttavia, il *citation index*, per essere significativo, richiede un orizzonte temporale lungo e sicuramente nel 2005-2006 non può essere usato per valutare la produzione del triennio 2001-2003. Inoltre, deve essere depurato dalle autocitazioni, che in alcuni settori, con numerosissimi coautori, possono essere tante. Poi l'area di appartenenza in cui i vari articoli vengono aggregati - al fine della costruzione da parte dell'ISI del *quotation index* - non si riferisce alle 14 + 6 macro-aree definite nell'esercizio del CIVR, bensì a micro-settori molto più dettagliati. L'IF è un criterio più rozzo, ma più rapido da utilizzare (sempre combinato con la *peer review*), sia nel senso che non chiede un *lag* temporale rispetto alla pubblicazione dell'articolo, sia perché sono facilmente ritrovabili gli *impact factor* delle 10-20 più autorevoli riviste per area, dove normalmente sono diffuse le ricerche migliori. Il che è tanto più una qualità quanto più si voglia assicurare, a chi si sottopone a giudizio, rapidità e ineccepibilità nella valutazione a costi relativamente contenuti, tenuto conto che sono necessari confronti interdisciplinari e interspaziali, dal momento che sia il *citation index* che l'*impact factor* sono mediamente molto differenziati fra discipline diverse. Mi spiego con un esempio. Secondo i dati del CIVR, la quota dei lavori di economia giudicati eccellenti dagli esperti è, dopo agraria e veterinaria, la minore (sotto il 20%, a fronte di una media nazionale del 30%). Se si guarda agli IF delle riviste dove gli eccellenti in economia italiani hanno pubblicato, rispetto ai massimi IF delle riviste di economia nel mondo e si confronta con l'equivalente in fisica, si capisce che l'insufficiente risultato degli economisti rispetto ai fisici italiani non dipende da una eccessiva severità dei valutatori CIVR della "scienza triste", bensì da una qualità dei prodotti economici mediamente meno all'altezza dei migliori mondiali, relativamente a quelli della fisica italiana. Contenere i costi e i tempi della valutazione, rimanendo affidabili, può essere la ragione principale per ricorrere a sistemi misti di *peer review*, combinata con indici bibliometrici.





Se sì, con quali formule o caratteristiche oggettive?

Nell'escogitare, dunque, una buona formula di valutazione, 3 caratteristiche oggettive sembrano particolarmente importanti.

Rapidità: se il giudizio viene tardi, come nella giustizia, non serve più né a correggere né a incentivare.

Trasparenza/Affidabilità: chi è valutato deve capire perché riceve quel giudizio, e perciò questo deve essere scritto, motivato, consultabile dall'interessato, contestabile da altri, verificabile nella sua coerenza interna ed esterna.

Indipendenza (neutralità/terzietà): il *policy-design* deve aiutare l'indipendenza della valutazione in 4 sensi diversi:

fra valutatori: per fare un passo avanti rispetto all'esperienza del CIVR, bisognerebbe che i due *referees* chiamati a dare una valutazione scritta per ognuno dei prodotti autoselezionati fossero scelti da due panelisti diversi, anziché dallo stesso panelista, come finora successo;

dei valutatori rispetto alle istituzioni di appartenenza: a questo mira la costituenda Agenzia (cui sembra pensare il Ministro Mussi), che chiederebbe ai membri del suo *board* una aspettativa obbligatoria dall'Istituzione di origine (nel CIVR, due di noi abbiamo già operato questa scelta nel passato biennio per tale motivo);

dei valutatori rispetto al Governo e al MIUR: bisogna separare le strategie *top-down* – le decisioni politiche e amministrative su quante risorse elargire alla ricerca, quali grandi rami privilegiare – dalle scelte *bottom-up* – fatte dai pari e riguardanti gli specifici lavori più innovativi, le particolari strade più promettenti, i singoli prodotti migliori -;

dei valutatori rispetto ai valutati: i codici etici di condotta adottati dal CIVR sono uno degli strumenti per eliminare conflitti di interesse, che comunque vanno evitati in ogni modo.

Infine, poiché il *policy-design* può creare perfette condizioni istituzionali di indipendenza, ma non può impedire che gli uomini prescelti per l'azione, rivelandosi mediocri, desiderino essere dipendenti per vocazione opportunistica, per insipienza, per insicurezza, o perché sono gregari dei potenti, è bene sapere che un'eccellente *governance* non è certo sufficiente; per la verità, essa non è nemmeno necessaria, come il caso dell'Istituto di Contabilità Nazionale nordamericano dimostra, dove i vertici

sono subordinati al Ministro delle Finanze, eppure di fatto sono molto più indipendenti dei nostrani, spesso viziati da "*servo encomio*", pronto a trasformarsi in "*codardo oltraggio*", nonostante appaiano protetti da pressioni esterne, in Enti dotati di terzietà. Da questo punto di vista, la differenza fra gli USA e l'Italia sta nel fatto – e non è piccolo – che lì le pressioni del Ministro non condizionano, perché qualora intendessero impedire l'emergere della verità, il ricorso ad una stampa libera, seria e autorevole metterebbe in pericolo la testa di quello, non la testa del vertice tecnico. In Italia, per rimanere integri, bisogna possedere una forte tempra morale, che è sempre rara; negli USA, basta avvalersi della minaccia credibile del quarto potere. Questo svolge al meglio il suo ruolo, non semplicemente facendo paura ai politici in nome di altri interessi – come avviene da noi -, ma in ragione di una deontologia professionale intransigente, responsabile e votata al rispetto della verità.

Se sì, con quali tipi di valutatori o caratteristiche soggettive?

L'onestà intellettuale e la competenza sono quindi le due caratteristiche soggettive principali per ottenere una valutazione corretta, come sopra ricordato. Ma di una terza c'è bisogno se il compito, come nel CIVR, consiste nel giudicare i prodotti di ricerca in base a 4 elementi (misurazione il grado): qualità, rilevanza, internazionalizzazione ovvero potenziale competitivo internazionale, ed anche originalità/innovazione.

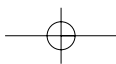
Perché allora è ovvio che i valutatori devono avere essi stessi un certo coraggio ed una propensione a percorrere vie nuove, anziché essere stanchi ripetitori.

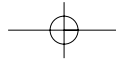
Su questi tre aspetti ho scritto recentemente un articolo su *Sissa News* e non vorrei ripetermi.

Qui desidero solo ricordare che onestà, competenza e coraggio sono doti rarissime e quasi inesistenti in probabilità in una sola persona adatta alla valutazione.

Perciò è la *governance* dei valutatori, nella sua composizione, che deve realizzare questa triade, fermo restando che poi è l'individuo prescelto che conta.

L'onestà intellettuale è più forte in chi non ha nulla da guadagnare o da perdere dai risultati della valutazione: perciò è opportuno (come nel





CIVR) che almeno 1/3 dei valutatori sia costituito da stranieri o da italiani all'estero.

La competenza, in un sistema scientifico che muta rapidamente, è più frequente fra i giovani di comprovata abilità (anche osservando il loro

record di quotation index e IF): ad essi va assegnato un altro terzo del dispositivo valutativo. Il coraggio di guardare lontano è più diffuso fra i vecchi, per natura presbiteri: la restante quota della *governance* deve essere per loro.

FIGURELLA KOSTORIS PADOA SCHIOPPA

Laurea alla Facoltà di Economia e Commercio dell'Università Bocconi di Milano con lode e dignità di stampa (1968). Master of Science in Economia alla Graduate School of Economics del Massachusetts Institute of Technology, Cambridge MA, "Passed with Honour" (1970).

Professore Ordinario di Economia Politica presso la Facoltà di Economia dell'Università di Roma "La Sapienza" (dal 1989).

Professore presso il College of Europe di Bruges (dal 2001), dopo essere stata Visiting Professor presso la Fondation Nationale des Sciences Politiques di Parigi e presso il Department of Economics della Columbia University; Visiting Scholar presso il Massachusetts Institute of Technology (MIT), presso l'Université Catholique de Louvain (CORE) e presso l'Università di Bruxelles; Jemolo Fellow presso il Nuffield College di Oxford; Research Fellow presso l'Italian Academy della Columbia University; Consulente per il Fondo Monetario Internazionale, Dipartimento per l'Europa.

Contatti:

Via in Publicolis, 43 int. 7
Tel/fax 06 68192450

00186 ROMA
E-mail: fiorella.kostoris@tin.it

