

Paolo Rossi

PROBLEMI E IPOTESI PER VALUTARE LA RICERCA



Lavorare per ipotesi

La ricerca in Italia, per motivi riconducibili in larga misura alla struttura e all'articolazione del sistema produttivo, è purtroppo confinata quasi totalmente al settore pubblico, le cui risorse sono peraltro scarse e per di più decrescenti.

Il posizionamento complessivamente buono della ricerca italiana (attestato da studi internazionali basati su metodologie oggettive e indipendenti) non è tuttavia omogeneo tra le differenti aree disciplinari né tantomeno può considerarsi assicurato a tempo indeterminato se non verranno messe in atto adeguate politiche di sostegno. Tali politiche dovranno necessariamente comportare una stretta combinazione tra una maggior disponibilità di risorse (umane e finanziarie) e un aumento di efficienza nel loro utilizzo. Premessa indispensabile a tale aumento di efficienza è certamente l'acquisizione di una reale capacità di valutazione *ex post* dei risultati, sulla cui base impostare politiche premiali e, ove necessario, disincentivanti.

L'aspirazione a metodologie valutative che possano a buon diritto definirsi "obiettive" è molto forte e non soltanto in Italia, ma la validità delle metodologie quantitative finora proposte (di natura prevalentemente bibliometrica) è tuttavia ancora oggetto di un vasto dibattito. Sarà quindi in questa fase necessario formulare ipotesi di lavoro, al momento evidentemente provvisorie, e metterle concretamente alla prova per misurarne la validità, sempre mantenendo una chiara consapevolezza del fatto che nel breve-medio periodo, a fronte di evidenze negative, occorrerà operare revisioni anche sostanziali della impostazione inizialmente adottata. In tal senso anche i documenti recentemente apparsi per opera di diversi soggetti istituzionali (in particolare il CUN e l'ANVUR) e finalizzati alla definizione di «Criteri e parametri per la valutazione», pur se finalizzati alla redazione di norme a carattere prescrittivo, devono essere considerati come provvisori, e gli effetti dell'eventuale adozione delle indicazioni in essi contenute dovranno essere attentamente monitorati per cogliere gli eventuali segnali di reazioni perniciose da parte del sistema della ricerca, quali lo sviluppo di comportamenti opportunistici o l'abbandono di linee di ricerca culturalmente e scientificamente importanti ma poco remunerative in termini di valutazione "obiettiva".

Le tipologie di valutazione

Prima ancora di elaborare proposte operative può essere utile cercare di identificare le principali opzioni metodologiche e i loro possibili domini di applicazione, notando che in contesti differenti possono risultare adeguate opzioni anche molto diverse tra loro. La varietà

delle situazioni cui può riferirsi la valutazione della ricerca può essere sommariamente classificata secondo i seguenti principali parametri:

- a) finalità della valutazione (finanziamenti, retribuzioni, reclutamento, carriera accademica)
- b) soggetti della valutazione (individui, dipartimenti, atenei, enti)
- c) metodologia valutativa (*peer review*, criteri bibliometrici, metodi misti)
- d) oggetti della valutazione (articoli, monografie, brevetti, altri prodotti di ricerca)
- e) grandi aree disciplinari (scientifico-tecnologica, scienze della vita, umanistico-sociale).

In merito alle finalità della valutazione una prima e fondamentale distinzione è volta a separare le azioni valutative finalizzate alla pura e semplice verifica di superamento di prefissati requisiti minimi di produttività scientifica individuale (quali ad esempio quelli richiesti per il conseguimento di abilitazioni o per la progressione stipendiale) dalle azioni che hanno invece come obiettivo quello di formulare graduatorie (*ranking*), in genere mediante attribuzione di punteggi (*rating*), al fine di reclutamento, avanzamento di carriera, finanziamento di progetti, ripartizione di risorse e più in generale con obiettivi di incentivazione e premialità.

In linea di principio il primo aspetto, pur denso di importanti conseguenze pratiche, sembrerebbe potersi prestare abbastanza facilmente alla definizione di criteri di valutazione basati su parametri di natura prevalentemente quantitativa, in quanto il carattere non comparativo del giudizio potrebbe ridurre la valutazione alla verifica del superamento di una soglia quantitativa relativa a uno o più indicatori di produttività scientifica, ancorché certamente sia gli indicatori che la soglia debbano essere scelti con la massima attenzione all'accuratezza e all'efficacia.

In pratica la questione tende a complicarsi sotto diversi profili. In primo luogo, quando si voglia associare a questo tipo di valutazione qualche obiettivo strategico (ad esempio l'innalzamento della qualità media del personale ricercatore), sembra diventare inevitabile associare in qualche modo i valori di soglia a valori medi attualmente presenti nel sistema (o in qualcuna delle sue componenti). Tale associazione non è tuttavia scevra da rischi, in quanto la determinazione "storica" di tali valori medi e in certi casi la loro stessa definizione "algebrica" possono dipendere in modo talvolta cruciale dalle differenti dinamiche e modalità organizzative presenti nei vari settori disciplinari. Solo a titolo di esempio, si consideri l'importanza che può avere la presenza di una struttura fortemente gerarchica della ricerca, che porterà all'innalzamento dei valori medi dei gruppi dirigenti fino a livelli difficilmente raggiungibili per chi di quei gruppi dirigenti non fa ancora parte, oppure la rilevanza di una corretta stima del grado di proprietà individuale del prodotto di una collaborazione in quei settori in cui la dimensione delle collaborazioni può variare da pochissime unità a molte centinaia, con gli ovvi effetti differenziati di produttività.

Non bisogna poi trascurare il fatto che la richiesta di superare valori medi attualmente esistenti può essere penalizzante nei settori in cui il livello è già coerente con gli *standard* internazionali (e quindi non richiede in realtà di essere sistematicamente superato, mentre sarebbe ragionevole attestare la soglia su valori inferiori, anche se non troppo, alla media), e viceversa nei settori in cui la qualità media attuale è troppo bassa un criterio legato alla media non permetterebbe comunque un rapido miglioramento della qualità.

Tutte queste considerazioni sembrano suggerire l'importanza di rifuggire dalle facili generalizzazioni, e anche qualora si decida di fare riferimento prevalente a criteri e parametri di natura quantitativa occorrerà calibrare le soglie caso per caso tenendo conto della specificità dei settori di ricerca, oltre che della maggiore o minore facilità di pervenire a una

determinazione empirica attendibile dei valori di parametri che non in tutte le aree sono a tutt'oggi adeguatamente censiti e resi reperibili. Anche il riferimento a basi di dati bibliometrici internazionali non è sempre garanzia di completezza o di piena attendibilità anche solo quantitativa.

Ma soprattutto, nei casi in cui una valutazione, anche non comparativa, abbia comunque l'obiettivo di effettuare una forte selezione, è imprescindibile l'intervento di valutatori esperti che siano in grado di associare con competenza al giudizio quantitativo un giudizio di merito qualitativo, che in ultima analisi resta l'unica garanzia di non esclusione di soggetti di grande valore che per i più svariati motivi (compresa la giovane età, e perfino la reale originalità della ricerca svolta) possano trovarsi a non superare soglie meramente quantitative. Si pone allora il problema non marginale della selezione di questi valutatori esperti. Il principio della coerenza con i parametri adottati per la valutazione dei candidati appare generalmente ragionevole, ma ancora una volta è necessaria una specifica attenzione alle differenze anche grandi che certamente esistono fra i differenti settori della ricerca. I rischi sono gli stessi già esposti in precedenza, e ad essi si aggiunge anche quello assai concreto di una restrizione delle possibili rose di "esperti" a numeri così ristretti da costituire vere e proprie caste di professionisti della valutazione, che potrebbero diventare, per banali dinamiche di gruppo, detentori di un potere di scelta e di veto in ultima analisi sproporzionato anche a fronte delle loro pur riconosciute competenze.

Se già la valutazione individuale non comparativa presenta, come abbiamo visto, numerosi aspetti problematici, la valutazione comparativa vera e propria richiede un'analisi ancor più sofisticata e impegnativa, che dovrà necessariamente fare attento e distinto riferimento a ognuno dei parametri generali inizialmente elencati.

Restando fin d'ora inteso che anche la natura degli oggetti della valutazione e l'area disciplinare di riferimento comportano distinzioni talvolta cruciali e richiedono analisi appropriate e proposte specifiche (come già in precedenza accennato), sembra tuttavia opportuno organizzare il nucleo della discussione sulla base di due soli grandi assi di riferimento, quello relativo ai soggetti della valutazione, che possono variare dal singolo studioso ai grandi numeri di un ateneo, e quello relativo alla metodologia, che può variare con una certa continuità dalla valutazione "calda" e puramente qualitativa (*peer review*) a quella "fredda" e prevalentemente quantitativa degli indicatori bibliometrici.

I criteri di valutazione bibliometrici

Per stabilire quale relazione debba esistere tra la tipologia (e dimensione) dei soggetti valutati e il peso relativo che nella valutazione dovrebbero avere le due differenti metodologie dovremo preliminarmente esaminare in dettaglio l'origine e il significato dei criteri bibliometrici. Notiamo che ogni criterio quantitativo trae il proprio fondamento concettuale da considerazioni di tipo statistico, sia che si tratti di un riferimento a valori medi (come nel caso dell'*Impact Factor* o di qualunque altro tipo di *ranking* di riviste o di sedi di pubblicazione) sia che si tratti di valori assoluti individuali (come per il numero delle pubblicazioni e delle citazioni, l'indice *h* e simili) la cui interpretazione dipende comunque dalla comparazione con un grande numero di casi simili, ovvero dalla posizione occupata all'interno di una distribuzione. Ma nelle distribuzioni caratterizzate da una probabilità che, in percentuale, decresce al crescere del valore del parametro misurato (le cosiddette distribuzioni prive di scala) il significato della media è assai poco pregnante (non potendosi spesso definire la varianza), ed è stato più volte dimostrato, mediante il calcolo delle probabilità, che il rischio che un articolo pubblicato su una rivista di minor rango sia migliore

di uno pubblicato su una rivista più “qualificata” non è in genere trascurabile. Sulla base di questa fondamentale premessa è facile argomentare che l'adozione di parametri bibliometrici (in modo esclusivo o prevalente) ai fini di una valutazione comparativa di individui è di per sé un errore concettuale, anche indipendentemente dalla qualità (spesso opinabile) del parametro stesso.

Dovrebbe quindi apparire chiaro che la valutazione comparativa della ricerca individuale, a qualunque fine essa sia effettuata, deve comunque fare necessariamente ricorso a un giudizio di merito sulla produzione scientifica, formulato da valutatori indipendenti e dotati della necessaria competenza disciplinare (*peer review*), che potranno servirsi anche dei parametri bibliometrici interpretandoli alla luce delle proprie esperienze e conoscenze (*informed peer review*). Il riferimento a parametri quantitativi da parte dei revisori, pur non potendo condizionare il giudizio qualitativo, è importante al fine di evitare che elementi di assoluta soggettività possano portare a valutazioni totalmente arbitrarie e a discrepanze gravi tra i giudizi espressi da soggetti differenti.

Viceversa, è altrettanto vero che, quando il numero dei soggetti, valutati in modo aggregato, è sufficiente a giustificare l'ipotesi che le fluttuazioni di qualità non possano incidere significativamente sull'interpretazione del dato quantitativo, le stesse leggi della statistica ci assicurano un grado di attendibilità dei risultati dell'analisi quantitativa che cresce naturalmente al crescere del numero dei valutati. La comparazione di differenti Dipartimenti relativi alla stessa disciplina, una volta opportunamente standardizzata per tener conto delle dimensioni dei Dipartimenti stessi, può essere quindi plausibilmente effettuata in modo prevalente sulla base di indicatori bibliometrici. Con opportune operazioni di normalizzazione tra discipline differenti, anch'esse possibili su base statistica, l'applicazione dei metodi bibliometrici di valutazione potrebbe costituire anche la base per ripartizioni di risorse almeno parzialmente ancorate al “merito” delle singole strutture all'interno di un Ateneo dato, e anche per la formulazione di un giudizio complessivo sui singoli Atenei.

L'uso degli indicatori bibliometrici per le valutazioni collettive e aggregate non può comunque ridursi a una mera misurazione della quantità di produzione scientifica, che inevitabilmente indurrebbe a comportamenti opportunistici, privilegiando una produzione intensiva, scarsamente meditata e di conseguenza dotata di scarso impatto sulla comunità scientifica nazionale e internazionale. Per questo motivo si è andata sviluppando in alcune comunità di ricerca, soprattutto nelle aree scientifico-tecnologica e bio-medica, la ricerca di criteri e di parametri anch'essi di natura quantitativa, ma volti a misurare il grado di interesse con cui le pubblicazioni scientifiche vengono accolte. Malgrado le numerose e anche facili critiche, il conteggio del numero delle citazioni ricevute dagli articoli su rivista sembra generalmente costituire la base di partenza per la stima di questi parametri e indicatori di “impatto scientifico”.

A partire dal conteggio delle citazioni individuali si è poi evoluto il concetto di *Impact Factor* (IF) di una rivista, che in sostanza consiste nella valutazione del numero medio annuo delle citazioni per articolo ricevute dagli articoli pubblicati sulla rivista stessa. Come già spiegato più sopra è tuttavia assai improprio convertire l'IF di una rivista in un giudizio sul singolo articolo, e quindi sulla qualità della produzione scientifica individuale. Ma esistono anche fondati motivi, più volte discussi nella letteratura sull'argomento, per ritenere che un'applicazione meccanica dell'IF non produca un'adeguata parametrizzazione della qualità scientifica neanche nel caso di valutazioni collettive, in quanto l'ampiezza dell'intervallo di

valori empirici dell'IF (da frazioni dell'unità a diverse decine) è del tutto sproporzionata all'effettiva differenza di qualità media degli articoli pubblicati, e risponde pesantemente a logiche di diffusione editoriale e alle importanti differenze esistenti nelle dimensioni e nelle modalità comunicative delle diverse aree e settori disciplinari, con discrepanze rilevanti anche tra sottogruppi appartenenti allo stesso settore.

Nel campo delle scienze umane e sociali la situazione è resa ancor più complessa dall'assenza di basi di dati sufficientemente ampie e adeguate per il conteggio delle citazioni, e dal ruolo rilevante della produzione monografica, per la quale tale conteggio è intrinsecamente più difficile, se non addirittura impossibile, almeno con gli attuali strumenti di analisi.

A questo proposito pare opportuno sottolineare qui l'assoluta importanza, ai fini della realizzazione di efficienti processi valutativi a livello nazionale, dell'attivazione (da troppo tempo soltanto annunciata) dell'Anagrafe delle Ricerche dei Professori e dei Ricercatori. Questo strumento, per il quale sono già stati largamente sviluppati gli strumenti tecnici e giuridici, permetterebbe non soltanto una valutazione sistematica di numerosi parametri anche in settori per i quali mancano altre basi di dati adeguate, ma renderebbe anche possibile, con opportuni e già previsti adattamenti, la creazione di un *repository*, ossia di un deposito elettronico contenente non soltanto i metadati (indicazioni bibliografiche) ma anche i veri e propri prodotti della ricerca in formato digitale, con una semplificazione radicale degli aspetti operativi del processo valutativo anche per le procedure volte al giudizio qualitativo che richiedono la concreta disponibilità e accessibilità dei prodotti. I problemi legati alla protezione di questi materiali e alla tutela del *copyright* e della proprietà intellettuale sono tecnicamente risolvibili e non dovrebbero quindi costituire un ostacolo all'avvio di una così importante (e tutto sommato non troppo difficile) realizzazione.

In ogni caso la definizione di criteri bibliometrici per la valutazione della ricerca richiede un ripensamento complessivo, con l'auspicabile obiettivo di identificare criteri che da un lato siano per quanto possibile omogenei tra le differenti discipline, e dall'altro non penalizzino artificialmente differenti tradizioni culturali e modalità di comunicazione scientifica.

Il ranking delle riviste

Un'indicazione che appare emergere in misura abbastanza largamente condivisa è quella di operare una prima sommaria distinzione riconducibile in pratica alla verifica della presenza di revisori oppure, in subordine, l'attestazione dell'esistenza di un Comitato Scientifico autorevole e terzo. Per quanto riguarda poi specificamente gli articoli su rivista dovrà essere valutata con attenzione l'ipotesi di una classificazione sintetica (e sufficientemente elastica) che identifichi con procedure ampiamente condivise, all'interno dell'insieme delle riviste ritenute scientifiche, differenti livelli di qualificazione, garantendo comunque un'equilibrata ripartizione delle riviste tra i suddetti livelli. Il "punteggio" qualitativo da attribuire ai diversi livelli non dovrebbe in ogni caso essere riferito a un intervallo di valori troppo ampio. La determinazione di questo *pool* di riviste dovrebbe comunque risultare da un processo interattivo, nel quale le comunità scientifiche, anche tramite le proprie associazioni, giochino un ruolo determinante, e nel quale i criteri di riferimento, oltre quelli già in precedenza definiti in relazione alla scientificità, siano l'impatto (nazionale e meglio ancora internazionale) della rivista, la presenza nelle principali banche dati internazionali, l'autorevolezza della direzione scientifica, l'affidabilità della gestione organizzativa.

Per quanto riguarda le monografie, il cui ruolo nella trasmissione dei risultati della ricerca è particolarmente importante nel campo delle scienze umane, sembra molto arduo immaginare criteri di classificazione “oggettiva” dell’impatto scientifico che si basino sulla collocazione editoriale o su un qualche tipo di *ranking* delle Case Editrici, una pratica peraltro di dubbia legalità e del tutto sconosciuta in tutti i Paesi che hanno affrontato il problema della valutazione della produzione scientifica. Qualora si intenda, in sede valutativa, graduare in qualche misura il valore delle opere pubblicate in volume sembra quindi ineludibile stabilire, anche nel caso di valutazione aggregata, una qualche forma di giudizio comparativo mediante *peer reviewing*, malgrado l’impegno straordinariamente più elevato insito in tale procedura.

Valutazione e incentivi

Resta dunque confermato che una valutazione “fredda” delle strutture è in linea di principio possibile, e per certi aspetti anche auspicabile (soprattutto in relazione a costi, tempi e sistematicità della valutazione stessa). Ma l’impegno valutativo sarebbe comunque sostanzialmente sterile qualora esso non fosse accompagnato da meccanismi capaci di tradurre le differenze di produttività scientifica così misurate in un effettivo aumento delle risorse per i più meritevoli, erogato a partire dal livello dipartimentale: meccanismi incentivanti o disincentivanti che operino soltanto al livello degli Atenei ben difficilmente otterrebbero effetti positivi sul comportamento delle singole aree e settori scientifici. Una strategia di incentivazione dipartimentale è anche preconditione per una concreta valorizzazione dei nuclei d’eccellenza esistenti anche in quelle realtà che, per ragioni storiche, territoriali e/o organizzative non fossero in grado di raggiungere nel loro complesso i livelli qualitativamente più elevati di produttività scientifica.

I nuovi metodi di valutazione

Le soluzioni proposte fino a questo punto sono comunque da considerarsi largamente contingenti. Se infatti proviamo a collocarci in una prospettiva temporale più ampia possiamo facilmente immaginare che nel medio periodo (ovvero nei prossimi 5-10 anni) l’evoluzione delle modalità di comunicazione, anche scientifica, associata alla diffusione e allo sviluppo di Internet e dell’editoria elettronica, e soprattutto delle moderne forme di interazione collettiva (*social network* e simili) identificate anche dalla locuzione *Web 2.0*, porterà a nuovi meccanismi valutativi, al momento ancora in embrione nel campo della ricerca, ma già efficaci in altri campi di attività. Ci riferiamo ad esempio all’idea del “controllo aperto” lanciata da *Nature*, e consistente nell’idea di “esporre” i manoscritti per un certo tempo su Internet, con la possibilità di essere commentati da tutti i lettori interessati, prima che il comitato editoriale decida di “pubblicarli” (probabilmente soltanto in forma elettronica), anche sulla base dei giudizi che sono stati espressi. Possiamo anche aspettarci sostanziali raffinamenti di quella forma di controllo *ex post* che è data dalle citazioni, se si supererà l’attuale meccanismo centralizzato (fondato quasi esclusivamente su basi di dati proprietarie) in favore di un’analisi a tutto campo mediante motori di ricerca (sul modello, per quanto certamente ancora molto rozzo, proposto da *Google Scholar*), che permetterebbe facilmente di estendere questo tipo di valutazione anche alle monografie, soprattutto se le forme di pubblicazione elettronica, come prevedibile, andranno generalizzandosi, auspicabilmente con modalità *open access*.

In quest’ottica è più che mai importante mantenere un atteggiamento aperto nei confronti dei meccanismi e delle regole per la valutazione, evitando irrigidimenti normativi e burocratici che rischierebbero di risultare superati in un breve arco di tempo, e quindi di frenare

pericolosamente il pieno dispiegarsi delle potenzialità offerte alla ricerca realmente creativa dalle nuove forme della comunicazione scientifica.