

# **Scuola Democratica**

*learning for democracy*

PEER-REVIEWED JOURNAL

**N. 5 nuova serie**

# La valutazione della ricerca in Italia

S/D)))

di Alberto Baccini

Nei libri di storia economica l'Italia è caratterizzata come un Paese *late comer*, che è arrivata cioè in ritardo al processo di industrializzazione di fine Ottocento. Lo è anche dal punto di vista della valutazione, visto che l'introduzione sistematica di processi di valutazione del sistema universitario e della ricerca arriva con qualche decennio di ritardo rispetto alle prime esperienze britanniche. L'ANVUR (Agenzia Nazionale di Valutazione del sistema Universitario e della Ricerca) è stata insediata meno di un anno fa dopo una lunga gestazione. E proprio in questi mesi sta faticosamente partendo la Valutazione della Qualità della Ricerca (VQR). Il fatto di arrivare in ritardo avrebbe potuto rappresentare un indubbio vantaggio, sia dal punto di vista del disegno istituzionale che della realizzazione dell'esercizio di valutazione. Avrebbe infatti potuto permettere di evitare gli errori commessi dagli altri e avrebbe potuto permettere di adottare soluzioni e standard messi a punto da altri. Purtroppo sia nel disegno dell'ANVUR che nella definizione del VQR si è deciso di intraprendere una via italiana che appare dettata più dal clima 'antibaronale' degli ultimi anni, che da volontà genuina di innovazione e modernizzazione.

Sul tema della valutazione la situazione è paradossale. A livello internazionale valutazione della ricerca, bibliometria e scientometria (*scientometrics*) sono ormai oggetto di discipline a tutti gli effetti con cattedre, riviste e società scientifiche dedicate. In Italia, gli echi del dibattito internazionale, che arrivano già attutiti, sono completamente coperti dalle grida di coloro che vogliono usare strumenti bibliometrici fai-da-te come una clava, con il dichiarato scopo di mettere in crisi le baronie; e da coloro che negano che valutazione e bibliometria abbiano qualsiasi significato e utilità. La sensazione complessiva che ricava chi abbia seguito il dibattito italiano (che si sta svolgendo in gran parte su [www.roars.it](http://www.roars.it)) è che l'assenza di una cultura diffusa della valutazione impedisca un approccio scientificamente corretto alla questione, in grado di tenere conto degli indubbi vantaggi dell'uso di strumenti quantitativi, ma anche dei loro limiti e carenze. Può pertanto essere utile ripercorre i fondamentali della discussione.



## 1. La qualità, l'importanza e l'impatto della ricerca

L'obiettivo di un processo di valutazione è ricavare in modo sistematico informazioni sulla qualità della ricerca. Il risultato di ogni processo di valutazione deve contenere informazioni sintetiche, espresse in forma verbale o numerica, che siano chiare e comprensibili anche per i non addetti ai lavori, in primo luogo i decisori politici.

Ogni processo di valutazione è basato, in ultima analisi, sul riconoscimento da parte di uno o più pari della qualità di un prodotto di ricerca. Tale processo assume molte forme diverse, che dipendono dalla nozione di qualità utilizzata, dai soggetti che effettuano la valutazione, e dalla 'forma letteraria' in cui è espressa la valutazione. La nozione di qualità della ricerca è scivolosa ed elusiva a causa della sua natura multidimensionale. Per fare chiarezza è utile separare nettamente alcune dimensioni rilevanti che spesso vengono confuse tra loro. In riferimento a un prodotto di ricerca è possibile in linea di principio valutarne separatamente 1) la qualità (interna); 2) l'importanza; 3) l'impatto e la diffusione.

- 1) La qualità (interna). Un prodotto di ricerca è di qualità se nella sua produzione sono stati rispettati i canoni prevalenti nella scienza in un certo momento del tempo: il metodo adottato, il rigore del ragionamento, la robustezza dei dati, la chiarezza dell'esposizione ecc. Il giudizio di qualità può essere espresso compiutamente solo dai membri della comunità informata dei pari che hanno chiaro il riferimento al modo prevalente in cui la ricerca viene condotta nel mondo. Gli elementi alla base del giudizio possono essere mantenuti impliciti, oppure esplicitati. Il giudizio di qualità non solo può variare nel tempo, perché nel tempo si modificano gli standard di riferimento, ma presenta anche elementi ineludibilmente soggettivi. Può essere espresso con un rapporto più o meno lungo e strutturato, come avviene, per esempio, nel giudizio del revisore su un articolo inviato per la pubblicazione a una rivista scientifica; o con una valutazione sintetica numerica rispetto a una scala definita, le stelle utilizzate nel REF britannico. Non è inutile notare che tutte le forme letterarie del giudizio di qualità sono soggette al problema dell'affidabilità del giudizio dei pari chiamati a esprimerlo; l'uso di una scala numerica non indica un giudizio più oggettivo solo perché espressa in forma quantitativa.
- 2) L'importanza per la comunità degli scienziati, ovvero la capacità di un lavoro di influenzare le ricerche condotte da altri, di aprire la strada a nuove vie per lo sviluppo della scienza. L'importanza di un contributo può essere pienamente apprezzata solo nel lungo periodo: sono innumerevoli gli esempi di risultati che si sono rivelati fondamentali solo dopo un lungo periodo di tempo dalla loro pubblicazione; a volte sono addirittura gli storici della scienza a stabilire una priorità dimenticata.

- 3) L'impatto o la diffusione nella comunità scientifica. L'impatto è il riconoscimento tributato a un contributo scientifico dalla comunità dei pari in un certo periodo. Esso si esplicita principalmente nella forma della citazione. A partire dall'idea che la citazione possa essere considerata come un indicatore dell'impatto di un contributo scientifico, Eugene Garfield sviluppò negli anni Sessanta le prime analisi quantitative delle citazioni, che rappresentano i primi passi della bibliometria contemporanea.

La distinzione tra qualità, importanza e impatto è molto utile perché non necessariamente un contributo scientifico gode simultaneamente di queste tre caratteristiche (Baccini, 2010). La qualità è condizione necessaria, ma non sufficiente perché un lavoro abbia importanza. Gli artigiani della scienza producono una messe di lavori di qualità, che rispettano cioè gli standard definiti dalla comunità scientifica, ma solo una minoranza di quei lavori si dimostrerà davvero importante ai fini dell'avanzamento della scienza. Il fatto che un saggio appaia, per esempio, su una rivista scientifica soggetta a revisione dei pari, significa che è un lavoro di qualità perché rispetta gli standard adottati in quel momento dalla comunità scientifica. Non necessariamente quel lavoro sarà riconosciuto come importante; né il fatto di essere pubblicato su una rivista garantisce che esso sarà citato e avrà impatto. Purtroppo molta letteratura tende a sovrapporre, più o meno intenzionalmente, qualità e impatto. Se il lettore avrà la pazienza di leggersi i criteri di valutazione adottati per il prossimo VQR dall'ANVUR, potrà facilmente notare che queste nozioni sono sistematicamente confuse tra loro, a differenza di quanto accade, per esempio nell'impostazione dell'analogo esercizio di valutazione britannico (REF).

Dunque la qualità di un prodotto di ricerca può essere stabilita solo con il ricorso alla revisione dei pari, e il suo impatto può essere approssimato attraverso l'osservazione delle citazioni che ha ricevuto in letteratura. In linea di principio gli indicatori bibliometrici e citazionali usati nelle procedure di valutazione non sono altro che strumenti quantitativi in grado di sintetizzare le informazioni di base su qualità e impatto della ricerca ricavate da revisione dei pari e citazioni. Da questo punto di vista contrapporre strumenti bibliometrici e revisione dei pari appare del tutto fuorviante. Questo vale sia per quanti ritengono che la revisione dei pari sia intrinsecamente superiore all'uso di strumenti quantitativi, sia per quanti, all'opposto, ritengono che le statistiche citazionali siano intrinsecamente più accurate. Il punto sostanziale è che l'adozione delle misure bibliometriche comporta in realtà problemi tecnici così significativi che senza la loro attenta considerazione c'è il concretissimo rischio di sostituire alla soggettività dei revisori, la soggettività nella costruzione e nell'interpretazione delle misure bibliometriche.





## 2. Tre problemi nell'uso degli indicatori bibliometrici

Può essere utile ricordare quali sono i principali problemi connessi all'uso di strumenti bibliometrici (per una trattazione estesa in italiano si rimanda a Baccini, 2010; e al classico Moed, 2005). Il primo problema riguarda la significatività dei dati sui quali sono costruiti gli indicatori citazionali. La qualità di ogni indicatore citazionale riflette la qualità dei dati di base su cui è costruito. Se i dati di base non sono accurati, o soffrono di rilevanti distorsioni allora anche l'indicatore sarà non accurato e distorto. Questo significa che l'uso degli indicatori richiede una attenzione fortissima alla qualità dei dati bibliografici e citazionali sottostanti, ovvero alla qualità dei database su cui sono costruiti.

Il secondo problema riguarda il livello di risoluzione dell'analisi (prodotto di ricerca, ricercatore, gruppo di ricerca, dipartimento ecc.). Tanto più il livello è disaggregato, tanto più difficile è interpretare univocamente il risultato di un indicatore bibliometrico. Si prenda ad esempio l'*h-index* (Hirsch, 2005), che caratterizza con un unico valore numerico la produzione scientifica di un ricercatore, tenendo conto sia del numero di articoli che del numero di citazioni ricevute: un ricercatore con  $h=10$  ha pubblicato 10 articoli ciascuno dei quali ha ricevuto almeno 10 citazioni. Consideriamo adesso tre scienziati ciascuno con  $h=10$ . Il primo ha scritto 10 articoli ciascuno dei quali ha ricevuto 10 citazioni; il secondo non solo ha scritto 10 articoli con 10 citazioni ciascuno, ma altri 90 articoli con 9 citazioni ciascuno; e il terzo ha scritto esattamente 10 articoli ciascuno con 100 citazioni. Tra questi tre scienziati di equivalente, probabilmente, c'è solo l'indice  $h$ . L'applicazione meccanica e pedissequa di un indicatore in un processo di valutazione rischia cioè di far perdere al valutatore la multidimensionalità della performance scientifica del ricercatore. A fronte di questo, l'applicazione di un ben costruito indicatore – per esempio l'Impact Factor o anche lo stesso indice  $h$  – per confrontare l'impatto nella comunità scientifica di alcune riviste in un determinato intervallo temporale, appare assai meno problematica.

Il terzo problema è di approccio generale all'uso degli indicatori bibliometrici, spesso presentati sotto forma di classifiche dei migliori ricercatori, dei migliori dipartimenti, delle migliori università ecc. In linea molto generale questo modo di procedere ha l'indubbio vantaggio di fornire indicazioni di facile lettura; ma si tratta di un approccio pre-scientifico e pre-statistico. Almeno dalla fine dell'Ottocento, gli scienziati che fanno uso di dati empirici quando calcolano dati statistici, per esempio medie di fenomeni, li accompagnano con indicatori della significatività e dispersione. Gli indicatori bibliometrici sono invece spesso usati in isolamento, come se avessero significato assoluto, senza associare loro nessuna misura della variabilità.

## Conclusioni

La valutazione basata sulla revisione diretta dei pari non è esente da errori e manipolazioni e questo non ne fa lo strumento salvifico immaginato dai più fervidi fautori dell'autonomia della cittadella della scienza. La valutazione condotta con metodi bibliometrici non è preferibile perché più oggettiva e impersonale. La scelta tra le due modalità di valutazione è da giustificarsi, molto più prosaicamente, su considerazioni relative ai costi e benefici a esse associati. Le procedure di valutazione basate sul giudizio dei pari sono molto costose, ma permettono di raggiungere una risoluzione del giudizio molto elevata. Questo le rende ineliminabili quando si tratta di valutare un singolo progetto di ricerca.

La valutazione condotta con metodi bibliometrici è una alternativa assai meno costosa, per questa ragione particolarmente utile quando devono essere valutati insiemi complessi di prodotti di ricerca, ricercatori o istituzioni. Il problema principale della valutazione condotta con indicatori bibliometrici è che il loro uso corretto e la comprensione del loro significato richiede la conoscenza di molti aspetti tecnici che nel dibattito accademico e di politica della ricerca sono spesso trascurati. Ogni scienziato ha le carte in regola per giudicare i suoi pari; da questo non segue però che possa utilizzare a tal fine gli strumenti bibliometrici senza averli prima ben studiati. Molti degli abusi della bibliometria traggono origine proprio dalla convinzione contraria.

Purtroppo la VQR è stata disegnata in modo contraddittorio, ipotizzando in modo inconsistente che un giudizio raffinato sui singoli prodotti di ricerca possa essere dato attraverso strumenti bibliometrici. Questo difetto del disegno sta spingendo l'ANVUR a sviluppare una, per così dire, originale bibliometria, che maschera dietro l'apparente scientificità dei numeri, l'arbitrarietà e l'infondatezza metodologica dei criteri. I risultati della VQR rifletteranno queste scelte e non saranno, purtroppo, credibili. Il ministro in carica, su cui non ricade né la responsabilità della nomina dell'ANVUR, né della scrittura dei criteri di base del VQR, potrebbe ancora porre rimedio alla situazione. Se non lo farà, il sistema della ricerca italiano avrà perso, oltre a risorse che avrebbero potuto essere impiegate altrimenti, un'altra occasione per capire quanta ricerca si fa in Italia e quali sono gli standard di qualità prevalenti.

## Bibliografia

- Baccini, A. (2010), *Valutare la ricerca. Uso e abuso degli strumenti bibliometrici*, Bologna, Il Mulino.
- Moed, H.F. (2005), *Citation Analysis in Research Evaluation*, Dordrecht, Springer.
- Hirsch, J.E. (2005), «An Index to Quantify an Individual's Scientific Research Output», *Arxiv-physics*, 0508025 v5: 5, <http://arxiv.org/abs/physics/0508025>.