

Stanno diventando sempre più diffusi gli esercizi nazionali di valutazione della ricerca, che coinvolgono università ed enti pubblici di ricerca. Tuttavia “l’esigenza della valutazione”, unanimemente condivisa a livello teorico, diviene problematica quando bisogna decidere con quali metodi realizzarla in pratica. Tradizionalmente il metodo più adottato è stato il cosiddetto peer-review, che prevede la valutazione da parte di esperti dei prodotti di ricerca sottomessi a valutazione dalle istituzioni. Recentemente, lo sviluppo delle tecniche bibliometriche, che si fondano sull’analisi quantitativa delle citazioni ricevute da articoli pubblicati su riviste scientifiche internazionali, ha indotto diversi governi a introdurre la bibliometria, limitatamente alle cosiddette scienze dure, a supporto o in sostituzione della tradizionale peer-review.

In Gran Bretagna, i tradizionali Research Assessment Exercises, puramente peer-review, lasceranno il posto al prossimo Research Evaluation Framework, di tipo informed peer-review, dove i panel potranno avvalersi anche di indicatori bibliometrici per formulare il loro giudizio sui prodotti di ricerca sottomessi a valutazione da parte delle università. In Italia la Valutazione Triennale della Ricerca del 2006, anch’essa peer-review, sarà sostituita presumibilmente dalla Valutazione della Qualità della Ricerca (VQR), che adotterebbe un metodo ibrido: sarebbero i singoli panel, in base alla specificità della propria area disciplinare, a decidere se avvalersi unicamente della peer-review, della bibliometria o di entrambe, per valutare un numero di prodotti comunque limitato per singolo ricercatore. In Australia, invece, l’Excellence in Research for Australia, lanciato nel giugno 2010, nelle scienze dure valuta l’intero output di ricerca delle istituzioni, unicamente attraverso indicatori bibliometrici.

I pro e i contro di peer-review e bibliometria sono stati sviscerati in letteratura. Per quanto attiene alla valutazione del singolo output scientifico, dagli studi condotti emerge che, per quanto non si possa affermare quale dei due metodi sia migliore, esiste comunque una correlazione significativa e forte tra gli esiti di una valutazione peer-review e quelli di un esercizio puramente bibliometrico.

I forti limiti della peer-review negli esercizi nazionali di valutazione emergono, invece, quando la si applica alla valutazione comparata di individui, gruppi e organizzazioni di ricerca. Tali limiti scaturiscono dai vincoli di budget e di tempo, che impongono alla peer-review su larga scala di restringere la valutazione a una quota soltanto dell’output totale di ciascuna istituzione di ricerca.

La prima conseguenza di ciò è che la valutazione comparata peer-review può avvenire unicamente lungo la dimensione della qualità, ma non della produttività dei soggetti valutati,

mentre è proprio la produttività l'indicatore per eccellenza dell'efficienza di qualsiasi sistema produttivo. La seconda è che i ranking finali dei soggetti valutati sono fortemente dipendenti dalla quota del prodotto valutato. La terza è che la selezione dei prodotti da sottoporre a valutazione può risultare inefficiente, a causa sia della difficoltà di individuare i prodotti realmente migliori, sia di eventuali comportamenti opportunistici di soggetti che potrebbero anteporre la visibilità dei propri prodotti al successo dell'intera organizzazione. Ciò avrebbe un effetto distorcente sui ranking finali e sulla loro capacità di rappresentare la qualità reale dei soggetti valutati. La quarta conseguenza è che le valutazioni peer-review non offrono alcuna informazione alle organizzazioni sulla performance individuale, che permetterebbe di allocare le risorse al loro interno in funzione del merito.

Con una quota limitata di prodotti valutati per ricercatore (tre sui sette anni in osservazione, stando alle anticipazioni relative al prossimo VQR italiano) è impossibile formulare giudizi di valore affidabili sui singoli scienziati. Se, infatti, l'obiettivo macro-economico è quello di allocare le risorse in maniera efficiente, non è sufficiente che i fondi siano attribuiti alle istituzioni in funzione del merito, ma è anche necessario che queste, a loro volta, li di-tribuiscono al loro interno secondo i medesimi criteri meritocratici. Per ultimo, i costi e i tempi di esecuzione della peer-review sono così ingenti da rendere la loro frequenza inadeguata per stimolare efficacemente il miglioramento del sistema ricerca.

Anche il metodo bibliometrico, basandosi sull'analisi citazionale, presenta delle limitazioni, legate in primis alla copertura delle basi dati bibliometriche in termini di prodotti scientifici censiti. Nelle scienze dure, comunque, il 95 per cento dei prodotti sottomessi dalle università al VTR risultava censito nella base dati bibliometrica Web of Science. Un altro limite è connesso al tempo necessario affinché le citazioni di una pubblicazione si accumulino e rappresentino quindi una variabile accurata del suo impatto. A parte forse le Scienze Matematiche, in altra occasione abbiamo potuto dimostrare che una finestra temporale di due anni è già sufficiente ad assicurare una valutazione accettabile dell'impatto. Considerando i pro e i contro di entrambe le metodologie, attraverso il loro confronto diretto, abbiamo quindi concluso che il metodo bibliometrico è di gran lunga preferibile al peer review in termini di accuratezza, robustezza, validità, funzionalità, costi e tempi di esecuzione.

In riferimento ai limiti sopra richiamati, abbiamo constatato, infatti, che, variando la quota dei prodotti valutati 8 volte, nell'area disciplinare di Scienze Fisiche, per esempio, solo 8 università su 50 non saltano di decile nel rank di performance. Confrontando poi la qualità bibliometrica dei prodotti presentati da ciascuna università al VTR rispetto a quelli presentabili, questa risultava inferiore alla mediana nel 29,6 per cento dei casi in Scienze Agrarie e Veterinarie, nel 26,5 per cento dei casi in Ingegneria Industriale e dell'Informazione, nel 24,8 per cento in Scienze Matematiche e Informatiche. Paradossalmente, se tutte le università avessero scelto i prodotti migliori, dato l'esiguo numero di prodotti richiesti (1 ogni 4 ricercatori in ruolo nell'area

disciplinare), adottando l'algoritmo di valutazione utilizzato dal VTR, un'alta percentuale di università sarebbe risultata al primo posto a pari merito: da un minimo del 50 per cento in Scienze Agrarie a un massimo del 96 per cento in Scienze Chimiche.

Gli effetti di tutte queste distorsioni sono sintetizzabili nel confronto finale della classifica VTR con quella bibliometrica della produttività di ricerca degli atenei: le università che presentano salti di quartile tra le due classifiche variano da un minimo del 53 per cento delle valutazioni nelle Scienze Chimiche a un massimo del 77 per cento nelle Scienze Fisiche.

Se tanti e tali sono i limiti della peer-review, è inevitabile chiedersi perché i governi insistano con tale metodo. La risposta risiede, in molti casi, nella difficoltà di applicare le tecniche bibliometriche su larga scala. Cercheremo di illustrare in maniera semplice e sintetica queste difficoltà, che concernono essenzialmente la classificazione disciplinare degli autori e delle pubblicazioni, nonché l'attribuzione delle pubblicazioni ai veri autori.

È noto che gli ambiti di speculazione scientifica sono caratterizzati dalla diversa intensità di pubblicazione, vuoi per il diverso grado di copertura settoriale delle basi dati bibliometriche, vuoi per le diverse funzioni di produzione nelle varie discipline: un ricercatore in fisica, per esempio, produce in media 2,3 volte il numero di pubblicazioni di un matematico. Da ciò discende la necessità di operare confronti di produttività tra ricercatori dello stesso settore disciplinare; ma mentre in Italia esiste una classificazione ufficiale degli universitari in settori scientifico-disciplinari, negli altri paesi questa non c'è. Inoltre, anche l'intensità di citazione varia da disciplina a disciplina: un articolo di matematica riceve in media, dopo otto anni, 4,4 citazioni; uno di biologia 16,3. Ma varia anche all'interno della medesima disciplina, tra un settore e un altro.

Da qui la necessità di eseguire, prima del confronto citazionale, la field-standardization, attraverso i seguenti passi: 1) classificare ciascuna pubblicazione nel settore di pertinenza, tra quelli indicizzati nella base dati bibliometrica sorgente; 2) calcolare la media o mediana delle citazioni ricevute da tutte le pubblicazioni dello stesso anno e settore; 3) dividere le citazioni di ciascuna pubblicazione per la media o mediana della distribuzione. Poiché un ricercatore pubblica in genere in ambiti scientifici diversi, per confrontare la produttività scientifica di due ricercatori, anche se afferenti al medesimo settore disciplinare, è necessario prima individuare tutta la loro produzione scientifica, standardizzarla e, quindi, procedere al confronto degli indicatori standardizzati. Da ciò discende che non è possibile confrontare la performance di ricerca di più organizzazioni se prima non si è misurata quella individuale dei ricercatori che vi afferiscono. Ma per misurare quest'ultima in maniera automatizzata occorre superare due ulteriori ostacoli: l'identificazione e omologazione delle varianti con cui gli autori di un articolo

indicano la propria affiliazione; la disambiguazione della reale identità di un autore, dati i problemi di omonimia tipici di popolazioni di ampie dimensioni.

Ciò richiede evidentemente un'approfondita conoscenza del sistema paese da valutare.

Per l'Italia, questi ostacoli sono stati affrontati nel laboratorio di ricerca che fa capo agli autori, in cui è stato messo a punto un sistema di supporto alla valutazione della ricerca (www.disp.uniro.ma2.it/laboratorioRTT/Testi/Altro/DSS_Archivio.zip

) su dati Web of Science. Tale sistema è, ad oggi, l'unico al mondo a fornire misure comparate di produttività standardizzata di singoli individui su scala nazionale: dieci università, un ente pubblico, tre fondazioni di ricerca e due governi regionali si sono già avvalsi del sistema. In altri paesi esperienze simili, condotte con criteri scientifici, sono limitate a valutazioni o di singole istituzioni di ricerca o di singola area disciplinare. Se in altri paesi, quindi, il ricorso al metodo peer-review negli esercizi nazionali di valutazione è inevitabile, a meno di richiedere a tutte le istituzioni di ricerca l'intera produzione scientifica come nel caso australiano, non lo è affatto in Italia e la scelta di tale metodo richiederebbe almeno una verifica da parte dell'Agenzia Nazionale di Valutazione (ANVUR), recentemente insediatasi.

Per quanto utili e desiderabili siano poi i confronti internazionali, dobbiamo accettare che lo stato dell'arte non consente di stilare classifiche attendibili della produttività di ricerca delle organizzazioni. Nonostante ciò, i media ci inondano in continuazione di classifiche di varia provenienza, che non hanno nulla di scientifico e si rivelano tanto inverosimili per l'Italia, quanto pericolose. Se, infatti, la teoria senza misurazioni è sterile, le misurazioni avulse dalla teoria possono essere diaboliche.

Francesco Coniglione, su "La Voce" del 19.11.2010, riportava la classifica delle università italiane per numero di piazzamenti nei maggiori otto ranking internazionali. In testa si trovavano le università Sapienza di Roma, Bologna, Pisa, Torino; mentre occorre arrivare quasi in fondo per incontrare le Scuole Superiori Normale di Pisa e Sissa di Trieste; della Scuola Superiore S. Anna di Pisa nessuna traccia. Quale persona di buon senso in Italia potrebbe ritenere plausibili questi risultati? La mancata rispondenza dei risultati alle attese dovrebbe indurre i più accorti a interrogarsi sulla bontà del metodo di valutazione.

Un semplice esame degli indicatori utilizzati e del relativo peso nel determinare il posizionamento finale, rivela che tutte queste classifiche dipendono più o meno fortemente dalla dimensione degli atenei. Nella più discutibile di tutte, l'Academic Ranking of World

Universities della Shanghai Jiao Tong University, oltre il 90 per cento del risultato di performance finale dipende dalla dimensione dell'ateneo. Anche laddove la dipendenza è meno forte, come nel caso dell'altrettanto noto Times Higher Education's World University Ranking, tutte queste classifiche sono comunque inficiate dalla totale assenza della standardizzazione per settore degli autori e delle citazioni (field-standardization).

Più recentemente, hanno trovato particolare seguito i ranking delle migliori organizzazioni di ricerca italiane stilati da Via Academy, associazione di studiosi italiani in UK. Il ranking proposto scaturisce dalla somma degli h-index dei ricercatori a esse affiliati, che risultano tra i top 500 italiani per h-index estratto da Google Scholar. Oltre alla dipendenza dalla dimensione (non a caso ai primi quattro posti troviamo le Università di Padova, Bologna, Milano e il CNR) e all'assenza di standardizzazione (l'h-index, per definizione, non contempla alcuna field-standardization), i ranking sono ulteriormente distorti dall'inadeguatezza della fonte, Google Scholar, e dalla limitata quota dei ricercatori valutati sul totale della popolazione.

Se di valutazione della ricerca si tratta, si auspica che i metodi impiegati siano il frutto della ricerca stessa e non dell'improvvisazione, italiana o straniera che sia.

(Fonte: G. Abramo, C. A. D'Angelo, technologyreview.it 19-08-2011)