



www.lavoce.info

[Scuola e Università](#)

CI VUOLE METODO PER VALUTARE LA RICERCA

di [Giovanni Abramo](#) 15.12.2009

Gli esercizi di valutazione della ricerca sono condotti in Italia con il sistema "informed peer-review". Che ha però molti limiti di precisione, affidabilità, robustezza, funzionalità, costi e tempi di realizzazione. Difetti che invece non si ritrovano nel sistema di supporto alla valutazione della ricerca che si basa sui dati dell'Osservatorio della ricerca pubblica. Ed è paradossale che si valuti la ricerca ignorando i progressi fatti nelle metodologie di valutazione.

Il primo esercizio nazionale di valutazione triennale della ricerca, il Vtr 2001-2003 condotto dal Civr, il Comitato di indirizzo per la valutazione della ricerca, ha diffuso tra le organizzazioni pubbliche la febbre della valutazione.

La legge 9 gennaio 2009, n. 1 con cui una parte del fondo ordinario delle università 2009 è stato assegnato agli atenei secondo indicatori legati tra l'altro agli esiti del Vtr; le ripetute dichiarazioni del ministro Mariastella Gelmini secondo cui una quota crescente del fondo sarà allocata nel futuro in funzione del merito e l'imminente lancio dell'esercizio quinquennale di valutazione Vqr 2004-2008 hanno scatenato una vera e propria pandemia. Più una pandemia si diffonde, si sa, meno le azioni tendono a essere ispirate dalla razionalità e più opportuno sarebbe il ricorso alla scienza medica piuttosto che al "fai da te". Una riflessione sullo stato dell'arte della scienza e tecnologia della valutazione della ricerca potrebbe rivelarsi quindi quanto mai utile in questo momento, sia per le istituzioni di governo sia per le organizzazioni di ricerca.

IL SISTEMA "INFORMED PEER-REVIEW"

Il sistema *informed peer-review* è un'evoluzione del *peer-review*, da cui si differenzia per il fatto che ai valutatori vengono fornite, ove disponibili, informazioni di natura **bibliometrica** (citazioni, impact factor) sui prodotti sottomessi a valutazione. I problemi di fondo della *peer-review* comunque rimangono.

Precisione: la *peer-review* prevede che un prodotto sia valutato da due esperti. È lecito chiedersi se due sole persone possano essere più precise nella valutazione rispetto all'intera comunità scientifica mondiale, attraverso la selezione della rivista prima e le citazioni poi. Ed è lecito chiedersi anche con quale criterio vengono selezionati i valutatori: bibliometrico, per caso? Chi verifica poi la veridicità della proprietà dei prodotti sottomessi? Nella prima esperienza australiana, del 1997, i prodotti sottomessi risultarono affetti dal 34 per cento di "errori", causando il 97 per cento di errori nelle risultanti classifiche.

Robustezza: la *peer-review* inevitabilmente impone una valutazione comparata basata su una sottopopolazione dell'output di ricerca. Questo pregiudica ogni misura di produttività che è l'indicatore di efficienza per antonomasia di qualsiasi sistema produttivo. Il precedente Vtr ha valutato un prodotto ogni quattro accademici nel triennio, ossia circa il **9 per cento** dell'intero output. Si può fondare una comparazione su un sottoinsieme così ristretto? Si ribatte affermando che quello che interessa è valutare e incentivare le **eccellenze**. Ma si può stabilire a priori quale sia la quota eccellente del prodotto totale? Ammesso e non concesso che sia il 9 per cento, dovrebbe essere diversa per ogni area disciplinare? **(1)** Per la diversa fertilità scientifica delle aree, un prodotto ogni quattro strutturati corrisponde a circa il 5 per cento dell'output totale in scienze fisiche e al 22 per cento in scienze agrarie e veterinarie. Quanto è robusta la classifica di performance che ne deriva? Un'analisi di sensibilità alla variazione della quota di prodotto valutato, mostra che, passando da un prodotto su quattro al 9 per cento fisso dell'output totale in ogni area (ossia a parità di numero di prodotti totali valutati), la **classifica per qualità** delle università varia notevolmente. **(2)** Nelle scienze fisiche, quaranta università su cinquanta cambierebbero posizione, con un salto massimo di quindici posizioni. Nelle scienze biologiche quarantacinque su quarantatré e ventidue posizioni. Variando la quota di prodotto valutato otto volte, tra il 4,6 e il 60 per cento. Nelle scienze fisiche, ad esempio, solo otto università su cinquanta ricadrebbero sempre nello stesso decile di ranking.

Affidabilità: si afferma che gli esercizi *peer-review* nazionali valutano la qualità della ricerca delle organizzazioni. Ne siamo proprio convinti? In realtà valutano la qualità dei prodotti sottomessi a valutazione, che non necessariamente rappresentano il meglio del prodotto di ricerca di un'organizzazione. A parte l'esercizio del potere di persone o gruppi che potrebbero anteporre la rappresentatività personale alla qualità intrinseca dei prodotti nei criteri di selezione interna (selezionare il *paper* del collega significherebbe ammettere che è migliore del proprio), esistono difficoltà tecniche oggettive nel comparare prodotti di ricerca in settori diversi. A titolo di esempio, si pensi alla difficoltà incontrata dagli atenei nello scegliere, nell'esercizio Vtr, nell'area delle scienze mediche, il migliore tra tre prodotti, rispettivamente di dermatologia, cardiologia e neuroscienze. O per un ricercatore, nell'esercizio Vqr, due prodotti tra tutti quelli da lui pubblicati, come spesso accade, in settori diversi. Nell'esercizio Vtr circa il 30 per cento delle pubblicazioni sottomesse a valutazione nelle scienze agrarie e veterinarie aveva un indice di qualità bibliometrico inferiore alla mediana dei prodotti realizzati; il 27 per cento in ingegneria industriale e dell'informazione e il 25 per cento in scienze matematiche e informatiche. **(3)**

Funzionalità: gli esercizi nazionali di valutazione della ricerca sono funzionali al perseguimento di diversi obiettivi di policy. Primo fra tutti, in Italia, un'allocazione efficiente delle risorse. In una prospettiva macroeconomica, affinché ciò si realizzi, non è sufficiente allocare le risorse alle organizzazioni di ricerca migliori, ma è anche necessario che queste a loro volta, indirizzino al loro interno le risorse ai ricercatori o gruppi di ricerca migliori. Sotto questo aspetto, gli esercizi *peer-review* nazionali non aiutano in alcun modo gli atenei, perché non danno loro alcuna informazione utile allo scopo. Sistemi di valutazione che misurino la performance comparata a livello di singoli ricercatori sarebbero senz'altro più funzionali.

Costi e tempi: i costi diretti degli esercizi *peer-review* sono molto alti e sono funzione della quota di output totale valutato. Nel Regno Unito il Rae 2008 è costato 12 milioni di sterline. Cinque volte maggiori sono stimati i costi indiretti (costo opportunità del tempo del personale di ricerca e amministrativo dedicato alla partecipazione al Rae) sopportati dalle organizzazioni sottoposte a valutazione. Anche i tempi sono lunghi, non meno di due anni per espletare l'intero esercizio. Il tutto ha come conseguenza una frequenza di cinque-sei anni degli esercizi, relativamente bassa rispetto a quella necessaria per incentivare efficacemente il miglioramento del sistema ricerca.

Se tanti e tali sono i **limiti** della *peer-review*, perché nel mondo si continua a utilizzarla? Va detto che in altri paesi la quota di prodotto valutato è senz'altro maggiore che in Italia e non viene fissata in funzione del numero di addetti: le organizzazioni sono libere di sottomettere tutti i prodotti che

ritengono eccellenti. Nonostante ciò il metodo di valutazione *peer-review* non può essere considerato soddisfacente. I sistemi bibliometrici dal canto loro non hanno potuto offrire una valida alternativa su larga scala, per la difficoltà di associare ciascuna pubblicazione al suo vero autore. A oggi, nessun paese al mondo dispone infatti di una base dati a livello nazionale della produzione scientifica dei singoli ricercatori. **(4)** Nessuno, tranne l'Italia.

IL SISTEMA SSV-ORP

La base dati Osservatorio della ricerca pubblica (Orp), derivata da *Web of Science* su licenza Thomson Reuters (già Isi) censisce tutte le pubblicazioni scientifiche (272mila) e *conference proceeding* (100mila) di organizzazioni pubbliche (354) e private localizzate in Italia, a partire dal 2001. A ciascun autore universitario sono attribuite tutte le sue pubblicazioni con un errore inferiore al 5 per cento. I relativi indici di qualità (citazioni e impact factor) sono normalizzati in funzione dell'intensità di citazione del settore cui afferisce la pubblicazione (e non l'autore). Il dataset è fortemente rappresentativo dell'intero output di ricerca: basti pensare che sono censiti in Orp il 95 per cento dei prodotti presentati al Vtr nelle aree disciplinari universitarie Adu 1-9 (le cosiddette *hard sciences*), in alcuni settori scientifico disciplinari delle scienze economiche e sociali e in psicologia, ossia per il 70 per cento dell'intera ricerca universitaria. A partire da Orp è stato realizzato un **sistema di supporto alla valutazione** della [ricerca Ssv-Orp](#), che consente di stilare **classifiche** per diversi indicatori di performance di tutti i ricercatori italiani per settori scientifico disciplinari, limitando le distorsioni dovute alla diversa fertilità degli ambiti di ricerca. Aggregando i dati, è possibile poi misurare il posizionamento di interi settori scientifico disciplinari e, pesando questi ultimi per dimensione, di aree disciplinari universitarie, atenei ed enti pubblici di ricerca. Infine, esprimendo la posizione nella classifica nazionale in rango percentile, è possibile comparare non solo la performance di settori scientifico disciplinari, aree disciplinari, atenei o enti di ricerca diversi, ma anche di raggruppamenti di ricerca non omogenei, all'interno dello stesso ateneo o ente di ricerca, quali gruppi di ricerca, dipartimenti o istituti. Tali misure possono essere effettuate anche normalizzando rispetto al livello di inquadramento del personale di ricerca (ordinario, associato, ricercatore) e le classifiche stilate per ciascun livello.

L'Ssv-Orp è assolutamente non invasivo, non richiedendo alcun input da parte delle organizzazioni sotto osservazione, con conseguenti **risparmi** di costi e tempi di esecuzione. Censendo la produzione scientifica dei singoli, permette di condurre valutazioni comparate delle organizzazioni in base al totale degli strutturati, ai soli produttivi o ai soli top-scientist. Dalle analisi delle *co-authorship*, è possibile anche valutare il tasso di internazionalizzazione della ricerca e di collaborazione pubblico-privato.

Nelle aree disciplinari ove risulta rappresentativo, la superiorità di Ssv-Orp sulla *peer-review* è evidente in ciascuna dimensione: per precisione e robustezza perché valuta oggettivamente la quasi totalità dell'output. Per affidabilità perché evita le distorsioni dovute alla selezione interna. Per funzionalità perché i ministeri cui afferiscono le istituzioni di ricerca e le istituzioni stesse avrebbero tutte le informazioni per procedere alle rispettive allocazioni delle risorse in maniera efficiente. Per i costi perché riduce drasticamente quelli diretti e elimina quelli indiretti, e per i tempi perché si svolge in un ordine di mesi. Il tutto rende inoltre possibile una maggiore frequenza di valutazione.

Una valutazione nazionale che utilizzi il sistema Ssv-Orp nelle discipline in cui la pubblicazione è rappresentativa del prodotto di ricerca e la *peer-review* nelle rimanenti, sarebbe non solo possibile, ma anche auspicabile. In caso contrario, il paradosso di valutare proprio la ricerca ignorando i progressi fatti nelle **metodologie di valutazione** sarebbe veramente eclatante. Alcune organizzazioni hanno già sperimentato Ssv-Orp. Al ministero e alle organizzazioni di ricerca spetta ora il compito di decidere se utilizzare la radiografia o la tomografia computerizzata, a un costo paradossalmente inferiore, per formulare la diagnosi e indicare le terapie più appropriate per il bene pubblico "ricerca".

(1) La quota eccellente del prodotto totale è del 30 per cento in Nuova Zelanda, del 50 per cento nel Regno Unito ed è stimata del 21 per cento nel prossimo Vqr.

(2) Abramo et al., 2009a.

(3) Abramo et al., 2009b.

(4) Van Raan 2008.