

VALUTAZIONE DELLA RICERCA IN MATEMATICA

Documento Commissione Scientifica UMI
approvato dall'Assemblea dei Soci UMI del 15 maggio 2010
pubblicato sul Notiziario di giugno-luglio 2010

Questo documento vuole fornire alcune riflessioni e indicazioni sulle pratiche di valutazione della ricerca in campo matematico e sul loro uso, riflessioni e indicazioni che possano essere utili in tutti gli ambiti in cui sia necessario effettuare tali valutazioni.

A) Valutazione dei prodotti della ricerca

I prodotti della ricerca matematica da sottoporre a valutazione sono principalmente pubblicazioni (articoli su rivista o volumi di *proceedings*, libri o capitoli di libri, curatele ed edizioni critiche), e in misura minore software e prodotti multimediali.

Per la valutazione di ciascun prodotto si possono utilizzare, singolarmente o combinate a seconda della situazione e dell'uso che s'intende fare del risultato, le due metodologie seguenti:

- a) *valutazione di contenuto* del prodotto, da parte dei responsabili della valutazione o utilizzando revisori esterni (*peer review*);
- b) *valutazione di contesto* del prodotto, per esempio tramite indicatori bibliometrici o esame della sede di pubblicazione.

Il pregio fondamentale della valutazione di contenuto è che è l'unica che permette, in via di principio, di verificare effettivamente la qualità del singolo prodotto, in quanto lo esamina direttamente; inoltre, per alcuni prodotti può essere l'unica disponibile. I difetti principali sono i seguenti:

- richiede un ingente investimento di risorse (sia di tempo sia di persone), e quindi è utilizzabile sistematicamente solo per azioni di valutazione sufficientemente importanti da giustificare un simile investimento;
- il risultato, anche non considerando comportamenti scorretti, dipende dai valutatori o revisori scelti, che devono dare garanzie di competenza e affidabilità ed evitare situazioni di auto-referenzialità o conflitto di interessi, per cui si pone il serio problema di chi debba

- scegliere i valutatori e i revisori, e con quali criteri;
- il metro di giudizio adottato dai singoli revisori non è uniforme, nonostante i tentativi fatti dai responsabili della valutazione, in quanto è inevitabile che revisori diversi abbiano sensibilità diverse, e questo rende difficile paragonare le risposte di due revisori, specie se di area diversa.

I pregi evidenti della valutazione di contesto sono che si basa su informazioni spesso facilmente accessibili e che in certi casi può essere effettuata direttamente dai valutatori senza bisogno di revisori esterni, per cui è più economica e veloce. Inoltre, la consultazione di una buona banca dati può dare a una persona competente un'idea immediata della rilevanza internazionale di un prodotto. Per esempio, la Matematica dispone della banca dati MathSciNet gestita dall'American Mathematical Society, molto completa e aggiornata per i prodotti di contenuto prevalentemente matematico (anche se più carente per sottoaree di confine).

D'altro canto, la valutazione di contesto soffre di molteplici difetti, che la rendono di utilizzo complesso in procedimenti importanti di valutazione¹. Tipicamente, la valutazione di contesto di un singolo prodotto si basa su un conteggio delle citazioni ricevute dal prodotto, e/o su una valutazione della sede di pubblicazione. I difetti principali dei metodi basati sul conteggio delle citazioni sono i seguenti:

- il numero di citazioni misura principalmente la popolarità di un singolo prodotto, che dipende solo in parte dalla sua qualità ma anche dalle dimensioni della comunità che lavora nel campo di ricerca in cui si situa il dato prodotto, e dalle consuetudini in fatto di citazioni di quella comunità. In particolare, questo rende estremamente difficile confrontare numeri di citazioni (e, a *fortiori*, tutti gli indicatori bibliometrici costruiti a partire da essi) di prodotti appartenenti ad aree diverse, e molto spesso anche di prodotti appartenenti a sottoaree diverse della stessa area;

¹Come segnalato in influenti documenti internazionali, quali per esempio il report “*Citation statistics*” redatto dall'*International Mathematical Union* (IMU) in cooperazione con l'*International Council of Industrial and Applied Mathematics* (ICIAM) e l'*Institute of Mathematical Statistics* (IMS).

- il numero delle citazioni dipende fortemente dall'età del prodotto, rendendo difficile confrontare numeri di citazioni di prodotti pubblicati in anni diversi²;
- ci sono problemi seri nella validazione e determinazione del numero di citazioni ricevute da un singolo articolo: *database* diversi forniscono citazioni diverse in quanto considerano fonti diverse, e la scelta stessa delle fonti è a volte effettuata in maniera non trasparente, è soggetta a errori, e, soprattutto per alcune sottoaree, può essere anche molto lacunosa;
- la scelta di un criterio di valutazione che dipende fortemente dalle dimensioni della comunità che si occupa di un dato argomento potrebbe tendere ad allontanare i giovani da argomenti di ricerca più innovativi o meno frequentati, indipendentemente dalla loro importanza o significatività.

Questi difetti suggeriscono che per la valutazione di contesto di un singolo prodotto, almeno nell'area Matematica, indicatori basati esclusivamente sul numero di citazioni siano da usare con estrema cautela, tenendo sempre conto delle consuetudini molto diverse delle varie sottoaree in fatto di citazioni, e della difficoltà di determinazione di questi indicatori per tutte le sottoaree di confine³. Un uso degli indicatori che non tenga conto di queste differenze darà inevitabilmente risultati non attendibili quando si confrontano prodotti di aree diverse, e spesso anche di sottoaree diverse. L'uso corretto degli indicatori derivati dal numero delle citazioni, che non dovrebbe mai essere esclusivo, richiede pertanto un'approfondita conoscenza di queste problematiche da parte del valutatore.

La sede di pubblicazione ha una ragionevole connessione, almeno in media, con la qualità del prodotto, in quanto gli articoli migliori tendono a essere inviati alle riviste che la comunità percepisce come migliori e, viceversa, le riviste che si ritengono migliori tendono a essere più selet-

²In Matematica, tipicamente articoli anche importanti ricevono poche citazioni nei primi anni dopo la pubblicazione, ma possono continuare ad accumulare citazioni per decine di anni.

³Intendiamo con "sottoaree di confine" tutte quelle sottoaree (quali la didattica della matematica, la storia della matematica, e tutte le sottoaree con vocazione applicativa) che spesso pubblicano in riviste non necessariamente matematiche.

tive nell'accettazione degli articoli. Bisogna però stare attenti a come viene valutata la qualità della sede di pubblicazione. L'uso di indicatori bibliometrici per tale valutazione ha l'indubbio pregio di essere semplice, ma non è privo di difetti:

- prima di tutto gli indicatori bibliometrici, essendo basati sul conteggio delle citazioni, soffrono degli stessi problemi segnalati sopra (dipendenza dalla comunità di riferimento, dipendenza dal tempo, possibile lacunosità delle fonti consultate);
- indicatori bibliometrici diversi forniscono graduatorie delle riviste sensibilmente diverse. Poiché l'accumulo delle citazioni nell'area Matematica è tipicamente lento, nel nostro campo sembra preferibile usare indicatori basati su citazioni ottenute in lassi di tempo più lunghi, quale il MCQ calcolato da MathSciNet; questo indicatore è però attendibile solo per riviste (e prodotti) prettamente matematici, ed è invece più deficitario per riviste (e prodotti) delle sottoaree di confine, per i quali conviene invece usare altri indicatori.

Occorre osservare che la gravità dei difetti segnalati è minore che nel caso degli indicatori relativi ai singoli prodotti, in quanto mediata sui grandi numeri, per cui un uso accorto e non esclusivo di tali indicatori per la valutazione della qualità della sede di pubblicazione non è da escludersi a priori. La Commissione scientifica dell'UMI si impegna a produrre nei prossimi mesi una classificazione delle riviste matematiche in fasce di qualità basata su criteri largamente condivisi dalla comunità matematica internazionale, e di rivedere periodicamente tale classificazione.

Vale inoltre la pena ricordare esplicitamente che la ricerca scientifica è, per sua natura, un'attività complessa e con molte sfaccettature, tutte importanti, per cui qualsiasi indicatore singolo potrà rifletterne solo alcuni aspetti, e solo una combinazione di metodi effettuata da esperti potrà dare a valutazioni sufficientemente accurate.

Queste considerazioni portano a suggerire per procedure di valutazione significative l'adozione di un **sistema misto, in cui i revisori consultino una pluralità di indicatori bibliometrici, e in cui l'uso degli indicatori bibliometrici non sia mai automatico ma sia temperato da una conoscenza da parte dei valutatori delle loro caratteristiche**

e limiti.

Riassumendo, per la valutazione della qualità dei prodotti della ricerca matematica si fanno le seguenti raccomandazioni principali:

- per procedure di valutazione con ricadute significative sulle strutture o sui singoli (quali valutazioni CIVR, valutazioni MIUR per l'attribuzione del FFO, e così via) la valutazione di contesto dev'essere sempre accompagnata da una valutazione di contenuto; per la valutazione di contenuto occorre porre particolare attenzione nella scelta dei revisori, che devono dare garanzie di competenza e di indipendenza; la valutazione di contesto deve basarsi principalmente sulla sede di pubblicazione, e deve fare particolare attenzione alle peculiarità dei prodotti situati in sottoaree di confine; infine, qualora le due valutazioni dessero risultati contrastanti, la valutazione complessiva dovrà essere adeguatamente motivata;
- per procedure di valutazione che considerano un grande numero di prodotti, con ricadute limitate sulle strutture o sui singoli, o in presenza di risorse limitate per lo svolgimento di tale procedure (per esempio, attribuzione di fondi di ricerca di ateneo) può essere usata, anche in maniera esclusiva, una valutazione di contesto basata sulla qualità della sede di pubblicazione o su indicatori bibliometrici, purché si tengano nel debito conto le diverse consuetudini delle varie sottoaree in fatto di citazioni e la difficoltà nella determinazione di questi indicatori per le sottoaree di confine;
- infine, in ogni caso la valutazione di qualità di un singolo prodotto deve dare un risultato compreso fra tre o quattro valori, non di più, in quanto non ci sono né gli elementi né la necessità di una classificazione più fine.

*B) Valutazione dei singoli ricercatori*⁴

La valutazione dei ricercatori è ottenuta da una combinazione delle valutazioni ottenute per i singoli prodotti della ricerca, con eventualmente altri elementi. L'esame delle procedure attualmente in uso o in sperimentazione

⁴Con "ricercatore" s'intende chiunque faccia ricerca, inclusi quindi professori universitari di qualsiasi livello, come pure ricercatori universitari e degli enti di ricerca.

suggerisce le seguenti indicazioni:

- nel campo matematico, per tutti i motivi indicati sopra, occorre porre particolare attenzione nella scelta di eventuali indicatori bibliometrici per la valutazione dei singoli ricercatori;
- i vari prodotti possono essere suddivisi in tipologie (per esempio, articoli su rivista, articoli su volumi di *proceedings*, libri, opere di ricerca/didattiche/divulgative, rilevanza nazionale/ internazionale, testo originale/edizione critica/curatele), e ciascuna tipologia può avere un peso diverso;
- è opportuno considerare, sia pure con pesi *ad hoc*, anche le opere didattiche o divulgative, perché insegnamento e disseminazione della cultura fanno parte dei compiti istituzionali dell'università, e perché la loro elaborazione richiede comunque competenze che sono legate alla ricerca;
- per evitare che la quantità possa sofferire alla qualità della produzione, può essere opportuno mettere un limite al numero di prodotti valutabili, limite eventualmente dipendente dalla tipologia di prodotto;
- la valutazione deve riguardare un arco temporale significativo, almeno triennale, e deve tener presente eventuali riduzioni nella produzione dovuti a impegni istituzionali e a congedi per maternità o altro motivo;
- esempi di altri elementi considerabili sono: la direzione di progetti di ricerca, l'aver tenuto seminari in congressi internazionali, la partecipazione all'*editorial board* di riviste internazionali; l'aver ricevuto riconoscimenti internazionali;
- il risultato finale della valutazione degli afferenti a una stessa struttura⁵ non dev'essere una graduatoria né un numero assoluto ma una suddivisione in tre o quattro fasce (senza alcuna limitazione a priori della distribuzione nelle varie fasce), perché una suddivisione più fine è eccessivamente dipendente da elementi casuali non stret-

⁵Stiamo supponendo che tutti gli afferenti alla struttura appartengano alla stessa area, o che comunque le procedure di valutazione dei prodotti della ricerca siano state calibrate in modo da dare risultati comparabili per i diversi afferenti alla struttura.

tamente legati alle qualità relative dei singoli.

C) Valutazione delle strutture

La valutazione delle strutture è ottenuta tramite una combinazione delle valutazioni ottenute dagli afferenti alla struttura assieme eventualmente ad altri elementi. Questa combinazione deve però permettere il confronto fra strutture diverse della stessa area, e fra strutture di aree diverse; per questo motivo, per quanto la valutazione in aree diverse possa basarsi su criteri diversi, questi criteri devono essere scelti in modo da poter dare risultati confrontabili. In altre parole, bisogna trovare un equilibrio fra due esigenze entrambe essenziali:

- i criteri e i parametri usati devono essere adattati alle singole aree, in modo da tenere presenti le caratteristiche distintive e gli usi consolidati di ciascuna area;
- i criteri e i parametri non possono essere eccessivamente diversi da un'area a un'altra, in quanto dev'essere possibile un confronto fra valori omogenei; in particolare, il confronto deve comunque avvenire fra strutture di dimensioni paragonabili.

Un tipico esempio di parametro che è strutturalmente disomogeneo fra le diverse aree è quello dei finanziamenti esterni, per almeno due motivi:

- aree diverse hanno possibilità di accesso a fondi esterni totalmente diversi (per esempio, i finanziamenti europei non coprono in ugual modo tutte le aree; alcune aree non hanno accesso a finanziamenti privati; e così via);
- la quantità di fondi necessari ad aree diverse è totalmente diverso (dipende, per esempio, dalla presenza o meno di grandi strumentazioni, e dalle dimensioni dei gruppi di ricerca).

D'altra parte, i finanziamenti esterni ottenuti da una struttura sono una ragionevole misura della qualità della struttura, se opportunamente considerati, per cui non possono essere scartati completamente.

L'esame delle pratiche attualmente in uso suggerisce le seguenti indicazioni:

- in caso di confronto fra strutture di aree diverse, i risultati della valutazione delle strutture *devono essere riportati su una scala nor-*

malizzata, e non essere valori assoluti, in modo da permettere un confronto omogeneo, e da evitare di favorire a priori strutture più grosse;

- ove sia disponibile, un parametro utile per il confronto fra strutture di aree diverse può essere il posizionamento ottenuto dalla struttura in valutazioni nazionali (quali la valutazione CIVR) rispetto ad altre strutture della stessa area e di dimensioni paragonabili;
- in caso sia usato un parametro relativo ai finanziamenti esterni ottenuti dalla struttura, possono essere considerati *solo* fondi a cui tutte le aree hanno accesso in ugual misura (per esempio, PRIN, ERC, qualche programma europeo ma non tutti), e *non* si deve tenere presente l'ammontare assoluto dei finanziamenti ricevuti;
- un criterio di valutazione molto grossolano che può essere utilizzato per la distribuzione di risorse relativamente poco significative è il rapporto fra il numero di afferenti attivi (cioè la cui valutazione abbia superato una certa soglia opportunamente normalizzata) e il numero totale degli afferenti alla struttura, tenendo presenti gli impegni istituzionali e i congedi per maternità o altro motivo;
- ove siano disponibili, è importante usare i risultati di valutazioni effettuate a livello nazionale anche per le distribuzioni di fondi interni agli atenei, in modo da premiare le strutture migliori (ed eventualmente quelle che più hanno migliorato il loro rendimento rispetto alla valutazione precedente).

D) Osservazioni conclusive

Le valutazioni nazionali (quali quelle messe in atto dal CIVR) devono chiaramente adottare il principio di paragonare a livello nazionale insieme di strutture dalle caratteristiche omogenee. Appare inoltre molto auspicabile che i risultati di tali valutazioni pesino in modo significativo sia nella distribuzione del FFO alle università sia nelle ripartizioni interne dei fondi dei singoli atenei. Si suggerisce inoltre un modello di ripartizione in cui vengano premiate sia le strutture migliori sia (benché in misura minore) le strutture che più hanno migliorato il loro rendimento rispetto alla valutazione precedente, in modo da incentivare ulteriormente comportamenti virtuosi e investimenti rivolti al miglioramento della qualità.

Si ricorda inoltre che i criteri di valutazione scelti influenzano inevitabilmente i comportamenti delle comunità sottoposte a valutazione. Per questo motivo, si auspica che la scelta dei criteri tenga esplicitamente presente le conseguenze che avrà, e che tale scelta sia effettuata e resa pubblica prima dell'esercizio di valutazione a cui si riferisce.

Infine, le procedure di valutazione finora messe in atto a livello di istituzioni pubbliche nazionali hanno avuto come seri limiti la loro episodicità e la loro variabilità. La messa in opera in tempi brevi di una Agenzia, pubblica e nazionale, di Valutazione della Ricerca che svolga con continuità, e quindi con la dovuta efficacia ed ampiezza, il lavoro di rilevazione e di valutazione appare quindi auspicabile e da tutti auspicata.