



**Università Commerciale Luigi Bocconi**  
Econpubblica  
Centre for Research on the Public Sector

## SHORT NOTES SERIES

**L'uso degli indici bibliometrici nella valutazione  
della ricerca: una nota**

*Carlo Devillanova*

**Short note n. 2**

May 2009

[www.econpubblica.unibocconi.it](http://www.econpubblica.unibocconi.it)

# L'uso degli indici bibliometrici nella valutazione della ricerca: una nota<sup>1</sup>

Carlo Devillanova, IAM ed Econpubblica, Università Bocconi

[carlo.devillanova@unibocconi.it](mailto:carlo.devillanova@unibocconi.it)

6 maggio 2009

1.

La valutazione della ricerca scientifica è un tema estremamente delicato. In un contesto di risorse scarse, è fondamentale ed opportuno che la ripartizione delle risorse fra istituzioni, progetti e persone tenga in debito conto la qualità della ricerca svolta; lo stesso vale per le progressioni individuali di carriera. Ne segue che i criteri adottati per valutare la ricerca possono indirizzare i comportamenti individuali ed avere profonde conseguenze sull'evoluzione della ricerca scientifica stessa.

Nella loro genericità, le precedenti affermazioni potrebbero essere ampiamente condivisibili. I problemi nascono al momento di definire la qualità della ricerca scientifica ed i criteri di misurazione della stessa.

1.1.

Negli ultimi anni si è diffusa l'opinione secondo la quale è auspicabile condensare in un unico indice numerico la qualità della ricerca scientifica. Il grande vantaggio di questo approccio è la sua semplicità, consentendo un'immediata comparazione degli oggetti della valutazione – siano essi individui, riviste, dipartimenti, centri di ricerca, università, ecc.

Un secondo presunto merito consisterebbe nella maggiore obiettività di un indice numerico rispetto ad altri meccanismi di valutazione. Questo argomento, particolarmente presente nel dibattito in corso nel nostro paese, presuppone una critica all'arbitrarietà dei sistemi di peer review in quanto fortemente manipolabili da parte di gruppi di potere del mondo accademico. Critica che ha senz'altro elementi di validità, ma che va confrontata alla luce dei limiti delle alternative proposte.

Infine, la quasi totalità degli indici proposti si basa su dati bibliometrici, primi fra tutti il numero di citazioni, che sono facilmente reperibili. Infatti, il terzo grande vantaggio (che mi pare evidente ed apprezzabile) di tale approccio è che rende l'esercizio di valutazione estremamente economico in termini di risorse finanziarie e di tempo. Probabilmente l'Impact Factor è il più noto fra questi indicatori ed è attualmente utilizzato da alcune prestigiose istituzioni per classificare la riviste scientifiche ed i prodotti di ricerca individuali.

1.2.

Recentemente, a livello internazionale è stato criticato l'uso degli indici bibliometrici come criterio (esclusivo) di valutazione della ricerca. In un paese, la Gran Bretagna, che può vantare una lunga e robusta tradizione di valutazione della ricerca, forti riserve sono state avanzate nei confronti di un uso meccanico dell'Impact Factor.

Giova a tal proposito citare alcuni passi del Research Assessment Exercise in UK, 2008, nei quali viene espressa una chiara censura nei confronti di criteri meccanicistici di valutazione della ricerca. Si legge nel RAE 2008: *“No panel will use journal impact factor as a proxy measure for assessing quality”* (Generic Statement n. 32); *“It is expected that most of the work submitted will be articles in refereed journals, books and research monographs, but the panel will not regard any form of output as necessarily being of higher or lower quality than others”* (Panel I criteria and working methods, p. 20; si veda anche p. 24 – Economics and Econometrics); *“The sub-panel confirms that the terms ‘international’ and ‘national’ refer to a quality standard, and not to the nature or scope of any particular research”* (Panel I criteria and working methods, sub-panel Economics and

---

<sup>1</sup> Questo scritto sviluppa un mio intervento ad un convegno sulla valutazione della ricerca svoltosi presso l'Università di Sassari nel marzo 2009.

Econometrics, p. 29); “Panels will consider the following individual circumstances (...): Engagement on long-term projects of significant scale and scope” (Generic Statement n. 32).

Sul piano metodologico, una lucida analisi delle statistiche bibliometriche e dell’uso improprio che si rischia di farne è svolta in Adler et al. (2008). Gli autori concludono che “*the sole reliance on citation data provides at best an incomplete and often shallow understanding of research—an understanding that is valid only when reinforced by other judgments. Numbers are not inherently superior to sound judgments*”, Adler et al. (2008), p. 2.

### 1.3.

Lo scopo di questa nota è squisitamente informativo. Credo sia utile e necessario richiamare le principali caratteristiche degli indici bibliometrici, per contribuire al dibattito in corso e per mettere in guardia contro un uso improprio degli stessi.

La mia discussione si concentra su due proprietà degli indici bibliometrici: la semplicità (o la perdita di informazione che la semplicità comporta) e l’oggettività. Più precisamente, nel secondo paragrafo discuto alcune questioni poste dalla creazione di classifiche di riviste. Il terzo paragrafo illustra i problemi connessi con la valutazione degli individui ed evidenzia come la pratica, sempre più adottata, di inferire la qualità di un articolo scientifico dalla sua sede di pubblicazione sia sostanzialmente scorretta. Il quarto paragrafo si concentra sul significato della misura del numero di citazioni, distinguendo fra qualità della ricerca scientifica ed impatto. Seguono alcune considerazioni conclusive.

## 2.

Il ruolo delle riviste scientifiche sembra aver subito una profonda evoluzione nel tempo, da mezzo di diffusione della conoscenza a strumento di certificazione della qualità della ricerca. Nel suo lucido editoriale al primo numero del AEJ Applied, 2009, Esther Duflo scrive che “*Papers are generally accessible and circulated long before they are published, either on the authors’ own Web page or through working paper series (...). Referred publications serve a central role in promotion decision, since they provide an independent assessment of a scholar’s production. (...) Those who do not need official stamps of approval publish much less in journals*”.

In quest’ottica, numerose istituzioni di molti paesi formalizzano classifiche delle riviste scientifiche, che vengono poi utilizzate ai fini della valutazione dei singoli ricercatori e dei gruppi di ricerca. Evidentemente, una più o meno esplicita graduatoria delle riviste scientifiche nelle specifiche aree disciplinari è esistita da sempre. A mio parere, la grande novità risiede nel tentativo di formalizzare queste classifiche basandole su indicatori numerici “semplici e oggettivi”, che consentano la comparabilità fra riviste di differenti aree disciplinari. Questa prassi ha due evidenti vantaggi: i criteri di classificazione delle riviste sono espliciti, rendendo il processo trasparente; la lista è aperta a integrazioni, una proprietà che deve essere accolta con estremo favore rispetto all’alternativa di liste bloccate<sup>2</sup>. Spesso, più o meno esplicitamente, gli indici bibliometrici (soprattutto l’Impact Factor) vengono utilizzati come criterio di ordinamento delle riviste. Di seguito evidenzio tre problemi legati alla costruzione delle liste.

### 2.1.

Un primo problema sollevato da questo approccio riguarda la comparabilità fra riviste afferenti a discipline diverse. In generale, gli indicatori numerici di qualità, in particolare il numero di citazioni, variano grandemente fra aree disciplinari per ragioni intrinseche all’oggetto della ricerca e per specifiche “culture” della citazione. Sul punto torno in seguito. Per il momento giova sottolineare che la comparabilità fra discipline è quantomeno problematica.

---

<sup>2</sup> Un esempio di lista chiusa in UK è la nota Diamon List.

## 2.2.

Con riferimento all'Impact Factor come criterio di classificazione delle riviste si pongono specifici problemi. Preliminarmente, occorre ricordare che l'Impact Factor di una rivista in un determinato anno misura il numero medio di citazioni degli articoli pubblicati su quella rivista nei due anni precedenti. L'indice è calcolato da Thomson Scientific.

Si tratta, quindi, di una statistica (la media) basata su un campione parziale. Almeno due sono le possibili distorsioni introdotte da questo metodo di calcolo (non mi soffermo, in questa sede, su aspetti di minore importanza quali alcune inconsistenze fra numeratore e denominatore nel calcolo dell'Impact Factor, o come esso dipenda del mese di pubblicazione degli articoli).

La prima distorsione riguarda la limitazione temporale ai due anni precedenti. Questa scelta rende estremamente problematica la comparazione fra aree disciplinari, in quanto la distribuzione temporale delle citazioni non è omogenea fra le stesse. Ad esempio, in medicina e biologia gli articoli ricevono la maggior parte delle citazioni entro due anni dalla loro pubblicazione, diversamente da quanto avviene in matematica e, notabilmente, economia. La figura 1, tratta da Adler et al (2008) e basata su dati Thomson Scientific, riporta l'anno di pubblicazione dei lavori citati nel 2003, distinguendo fra quattro aree disciplinari. Il periodo evidenziato dal rettangolo è quello che contribuisce al calcolo dell'Impact Factor. La figura dovrebbe mostrare con evidenza quanto appena detto.

Si noti, incidentalmente, che la media sul biennio precedente mostra una notevole variabilità nel tempo, soprattutto per le riviste con ridotto numero di articoli pubblicati ogni anno. Si noti anche che l'Impact Factor, per costruzione, recepisce con ritardo le variazioni nel numero di citazioni delle riviste; questo ritardo è evidente e grave in specifiche circostanze. Tutto ciò implica che i ranking di riviste basati sull'Impact Factor variano (con ritardo) nel tempo. Questa probabilmente non è una proprietà desiderabile in un meccanismo di valutazione della ricerca, soprattutto nella misura in cui ad esso sono associate anche finalità di indirizzo dei comportamenti individuali. Affinché i meccanismi di incentivazione della ricerca indirizzino in maniera virtuosa i comportamenti individuali e/o siano efficientemente utilizzati, devono essere noti con adeguato anticipo e relativamente stabili nel tempo.

La seconda possibile distorsione è introdotta dalla selezione del campione di riferimento, ovvero delle riviste considerate nel calcolo dell'Impact Factor. E' bene sapere che il campione include solo quelle riviste scientifiche selezionate da Thomson Scientific. Il grado di copertura differisce fra discipline, ma è comunque parziale. Le monografie scientifiche non hanno Impact Factor così come, evidentemente, le riviste di recente fondazione. In tal senso è implicita in queste classifiche una discriminazione fra prodotti di ricerca, contrariamente a quanto auspicato, ad esempio, nel Research Assessment Exercise 2008 citato in precedenza.

## 2.3.

Spesso i ranking raggruppano le riviste scientifiche in fasce di qualità omogenea, procedura che, in una certa misura, riduce i problemi legati alla variabilità temporale della qualità delle riviste. Tuttavia questa procedura è arbitraria e, per molti versi, opinabile: a mia conoscenza nessun indice di qualità delle riviste, in particolare l'Impact Factor, evidenzia particolari discontinuità. In altri termini, il raggruppamento in fasce tratta in maniera omogenea riviste sostanzialmente differenti rispetto a qualsiasi indice di qualità considerato e, al contrario, introduce artificiose discontinuità fra le riviste che si collocano agli estremi delle fasce stesse.

In verità, sono state sviluppate interessanti procedure di normalizzazione dell'Impact Factor per area disciplinare. Queste metodologie consentono di eliminare le suddette discontinuità, affrontando al contempo il problema di comparabilità delle discipline accennato in precedenza. Un aspetto delicato nell'applicazione di queste procedure riguarda la definizione dell'area disciplinare di riferimento, da utilizzare per standardizzare le misure numeriche di qualità. La scelta di un riferimento ampio (ad esempio, scienze sociali) rischia di riproporre a livello di singole aree sottodisciplinari il medesimo problema. E' noto, infatti, che anche all'interno di aree

apparentemente omogenee, la comunità scientifica di riferimento (e quindi la probabilità di pubblicare ed essere citati) e la cultura delle citazioni può differire grandemente. Al contrario, la definizione di piccole aree disciplinari pone problemi di comportamenti strategici non diversi da quelli che possono affliggere i processi di peer review.

La soluzione a questo problema non è semplice. Credo tuttavia che sia utile evidenziarlo per evitare gli opposti esiti dell'annichilimento di specifiche aree di ricerca o la loro legittimazione sulla base di indicatori di dubbia oggettività.

3.

Come già notato, in ultima analisi l'oggetto della valutazione sono gli individui, le istituzioni a cui appartengono, i progetti che intendono perseguire, ecc. A tal fine, gli indici bibliometrici possono essere utilizzati direttamente (calcolando, per esempio, il numero di citazioni di un determinato articolo o il numero di citazioni medie ottenute da un autore) o indirettamente (inferendo dalla qualità della sede di pubblicazione la qualità del prodotto della ricerca). Quest'ultima pratica è ampiamente diffusa, anche se si rivela assai problematica.

3.1.

La costruzione di ranking di riviste è funzionale ad un processo di valutazione indiretto, nel quale, cioè, al singolo articolo è associata la qualità media (l'Impact Factor) della rivista sulla quale è pubblicato. Se la rivista è buona l'articolo è buono.

La semplificazione implicita in questo tipo di inferenza dovrebbe essere evidente anche a chi non abbia particolari rudimenti di statistica. E' vero che in media il ragionamento è corretto, o tautologico, se si preferisce. Però è altrettanto vero che la probabilità che un singolo articolo sia di qualità pari a quella della rivista dipende crucialmente dalla distribuzione della qualità degli articoli nella rivista stessa. E' facile costruire esempi di riviste con Impact Factor identico, ma per le quali la probabilità che un singolo articolo, selezionato aleatoriamente dalle due riviste, venga citato lo stesso numero di volte è assai differente.

Il fatto è che la media è una misura cruda, che non coglie molti aspetti rilevanti per stabilire una valida inferenza statistica fra la qualità della rivista (comunque misurata) e la qualità dell'articolo ivi pubblicato. E' vero che il processo di referaggio delle riviste, unito al processo di autoselezione dei ricercatori nello scegliere la rivista cui sottoporre il proprio manoscritto, generano una certa omogeneità nella qualità degli articoli pubblicati su una data rivista. Tuttavia, è ben noto che tale qualità non è mai uniformemente distribuita. I dati sulle statistiche bibliometriche sono eloquenti a questo proposito e confermano una distribuzione estremamente polarizzata delle citazioni fra articoli, con una mediana pari a zero per la quasi totalità delle riviste. Per esempio, utilizzando delle riviste specializzate in campo biomedico, Seglen (1997) mostra *"the uneven contribution of the various articles to the journal impact (...): the cumulative curve shows that the most cited 15% of the articles account for 50% of the citations, and the most cited 50% of the articles account for 90% of the citations. In other words, the most cited half of the articles are cited, on average, 10 times as often as the least cited half. Assigning the same score (the journal impact factor) to all articles masks this tremendous difference—which is the exact opposite of what an evaluation is meant to achieve."*, p. 318.

Questo è causa due errori di segno opposto: l'attribuzione ad un articolo (o al suo autore) di una valutazione inferiore alla sua reale "qualità", perché non ha trovato un'adeguata collocazione editoriale; l'attribuzione ad un articolo (o al suo autore) di una valutazione superiore alla sua reale "qualità", se esso si posiziona nell'estremo inferiore della distribuzione della qualità degli articoli nella rivista sulla quale è stato pubblicato.

Ribadisco che il problema qui evidenziato non dipende dal criterio utilizzato per classificare le riviste ed il riferimento all'Impact Factor è conveniente al puro scopo esemplificativo.

### 3.2.

L'alternativa è quella di costruire direttamente un fattore d'impatto per la pubblicazione o l'autore. Per l'autore, probabilmente il più noto degli indici proposti è il fattore  $h$ , dato dal numero  $h$  di pubblicazioni di un determinato ricercatore che hanno ricevuto almeno  $h$  citazioni. In alternativa si può usare il numero medio di citazioni per pubblicazione, correggere l'indice  $h$  per l'anzianità scientifica del ricercatore, ecc. Tutte queste, ed altre, proposte si caratterizzano per il voler condensare in un unico indice le informazioni fornite dai dati bibliometrici dei singoli ricercatori. La giustificazione, ancora una volta, è legata alla semplicità di operare comparazioni interpersonali sulla base di valori numerici dell'indice.

La prima osservazione doverosa è che tale opera di semplificazione comporta perdite di informazione che rischiano di compromettere la supposta oggettività di questo tipo di comparazione fra ricercatori. Il problema è assolutamente evidente nel caso di numero medio di citazioni per articolo, non tenendo conto, questo dato, della numerosità della produzione scientifica. Ma anche il fattore  $h$  comporta una drastica perdita d'informazione, riducendo ad un singolo numero distribuzioni complesse. Si pensi, ad esempio, a tre ricercatori: il primo ha 20 pubblicazioni, ciascuna delle quali è stata citata 20 volte; il secondo 20 pubblicazioni, citate altrettante volte, e 200 ulteriori pubblicazioni citate "solo" 19 volte; il terzo ha 20 pubblicazioni, ciascuna delle quali è stata citata 10.000 volte. Il fattore  $h$  è identico nei tre casi considerati!

E' evidente che la semplicità di poter usare un singolo indice numerico ha i suoi vantaggi, ma i costi in termini di perdita di informazioni devono essere noti ed attentamente valutati.

La seconda considerazione riguarda la dubbia accuratezza di alcune delle basi di dati che alimentano queste statistiche. Chiunque può verificare, ad esempio, che l'indice  $h$  calcolato utilizzando il noto programma Publish or Perish viene influenzato dalla selezione o meno di discipline non attinenti ai propri ambiti di ricerca. Sono anche particolarmente frequenti gli errori causati da casi di omonimia, accessibilità su internet delle pubblicazioni, ecc. Delle distorsioni del campione utilizzato da Thomson Scientific per calcolare l'Impact Factor si è già detto.

### 4.

Uno degli aspetti più delicati nell'uso e nella diffusione degli indici bibliometrici riguarda la consapevolezza di cosa si sta esattamente misurando.

Fino a questo momento ho implicitamente ipotizzato che il numero di citazioni colga adeguatamente la qualità della ricerca scientifica. Credo che pochi siano disposti a sottoscrivere una tale affermazione. La stessa Thomson Reuters nel proprio sito web suggerisce di utilizzare questi indici per valutare la produttività e l'influenza dei singoli ricercatori, non la qualità della loro ricerca. Piuttosto, la "*citation analysis*" può essere utile come "*supplement to peer review and other forms of expert evaluation*" (thomsonreuters.com). Condivido appieno quest'ultima affermazione e sono fermamente convinto che un uso accorto dell'Impact Factor o di altri indici bibliometrici sia di grande supporto agli esercizi di valutazione della ricerca e, proprio per la loro economicità ed immediatezza, in molti casi desiderabile. Penso, ad esempio, ad un supplemento d'indagine conoscitiva che dovrebbe scattare in presenza di una significativa discrepanza fra la valutazione basata sugli indici bibliometrici ed altre forme di valutazione.

Detto questo, vedo elevato il rischio che un impiego acritico di indicatori numerici finisca con l'attribuire loro informazioni e virtù che non hanno: purtroppo, sono assai diffusi atteggiamenti che oserei definire feticistici nei confronti dei numeri.

Un uso accorto degli indici bibliometrici richiede, a mio parere, una seria riflessione su almeno tre questioni, alle quali accenno brevemente di seguito, ma la cui portata mi sembra travalichi abbondantemente la profondità di analisi di questa breve nota: cosa determina il numero di citazioni di un articolo; quale nesso esiste tra impatto di un articolo e qualità, intesa come influenza intellettuale dell'idea in esso contenuta; la stabilità nel tempo di tale nesso, dovuta alla potenziale manipolabilità del metro di valutazione.

#### 4.1.

La prima considerazione attiene alle determinanti della frequenza con cui un determinato articolo/autore viene citato. Spesso colgo negli interventi una sorta di “parabola del mercato” applicata alle citazioni scientifiche: se il mercato dei fruitori della ricerca scientifica (ricercatori in primis) cita un determinato articolo/autore ciò ne segnala l’influenza. *“The concept behind citation indexing is fundamentally simple. By recognizing that the value of information is determined by those who use it, what better way to measure the quality of the work than by measuring the impact it makes on the community at large. The widest possible population within the scholarly community (i.e. anyone who uses or cites the source material) determines the influence or impact of the idea and its originator on our body of knowledge.”* (thomsonreuters.com). In questa parabola, la domanda (di un’idea e del suo autore) determina il numero delle citazioni ed il valore dell’Impact Factor.

A proposito, faccio solo notare che le condizioni che determinano il (buon) funzionamento di un mercato sono numerosissime. Vale richiamare, solo a titolo d’esempio, i tanti problemi sollevati dalla presenza di asimmetrie informative, questione che mi pare di assoluto rilievo nel campo della ricerca scientifica di frontiera. Rilevanti sono anche considerazioni legate all’accessibilità dei prodotti di ricerca, che a sua volta dipende dalle tecnologie di diffusione utilizzate (si pensi ai prodotti presenti su database in formato elettronico) e, molto importante, la lingua di pubblicazione. Infine, non sottovaluterei affatto la struttura (non propriamente concorrenziale) del mercato dell’editoria scientifica.

Trovo sorprendente come nella mia disciplina (l’economia), che vanta una lunga tradizione di analisi dei cosiddetti fallimenti di mercato, questi problemi vengano spesso rimossi.

Un ulteriore aspetto merita di essere menzionato: un’importante determinante della probabilità con cui un articolo viene citato, a parità di ogni altra circostanza, è dato dal numero di potenziali fruitori dell’idea in esso contenuta. Questo crea importanti differenze fra aree disciplinari, che si sommano alle considerazioni fatte in precedenza e che rendono estremamente problematico l’uso degli indici bibliometrici per comparare riviste e autori di ambiti disciplinari diversi. Il tentativo di risolvere il problema attraverso le procedure di standardizzazione per area disciplinare presentano i limiti già discussi per le riviste e non sono di immediata applicazione ai singoli ricercatori. Il problema si presenta anche all’interno di aree disciplinari apparentemente omogenee, per specifici indirizzi di ricerca. Anche se si accetta la parabola del mercato, l’uso indiscriminato degli indici bibliometrici ha l’effetto di penalizzare i filoni di ricerca meno diffusi perché, ad esempio, più recenti o non inseriti nel main stream. Il problema del pluralismo è particolarmente importante e sentito nelle scienze sociali ed in economia in particolare (si legga Artoni, 2007).

#### 4.2.

Il secondo aspetto riguarda la necessaria distinzione fra impatto di un articolo ed influenza dell’idea in esso sviluppata. Un tentativo, probabilmente non esaustivo, delle ragioni che inducono alla citazione di un determinato articolo/autore include: citazioni che esprimono un debito intellettuale nei confronti dell’opera citata; citazioni utilizzate come artificio retorico; citazioni “negative”, quali, ad esempio, la confutazione di una determinata idea/risultato; citazioni di amici/conoscenti. Ultimamente, anche a causa dell’uso che viene sempre più spesso fatto degli indici bibliometrici, in alcune discipline si sta anche diffondendo il fenomeno delle citazioni introdotte per lusingare l’editore della rivista o il presunto referee<sup>3</sup>.

---

<sup>3</sup> Trovo pertinente e interessante quanto scritto dall’Editore di Economic Inquiry: “Journal time to publication lags have become embarrassing. Many authors have 5 year submission-to-print stories. More insidious, in my view, is the gradual morphing of the referees from evaluators to anonymous co-authors. Referees request increasingly extensive revisions. Usually these represent improvements, but the process takes a lot of time and effort, and the end result is often worse owing to its committee-design. Authors, knowing referees will make them rewrite the paper, are sometimes sloppy with the submission. This feedback loop - submitting a sloppy paper since referees will require rewriting combined with a

Evidentemente, solo il primo tipo di motivazione definisce un debito intellettuale nei confronti dell'opera citata. In tutti gli altri casi il nesso fra citazione ed impatto di un'idea e del suo autore è quantomeno labile. Inoltre, può darsi anche il caso contrario in cui il debito intellettuale nei confronti di un'idea non si traduca nella citazione dell'articolo che l'ha originata, perché la stessa è stata incorporata nel lavoro di altri autori<sup>4</sup>. Si noti, incidentalmente, che gli articoli di rassegna della letteratura – che in senso stretto non rappresentano progressi della disciplina – ricevono in media più citazioni.

Le due precedenti considerazioni portano ad operare una netta distinzione netta fra impatto e qualità.

#### 4.3.

Un argomento spesso addotto a favore di singoli indici numerici di valutazione della ricerca (bibliometrici o di altro tipo) è che essi hanno un'elevata correlazione (diversa fra discipline) con l'esito di più classici processi di peer review, ben più costosi in termini di tempo e risorse economiche. Un punto su cui, tuttavia, non mi pare si ponga sufficiente attenzione è che questa correlazione rischia di essere logorata da comportamenti strategici nel momento in cui tali indici venissero a rappresentare l'unico metro di valutazione della ricerca.

In altre parole, i meccanismi di ripartizione delle risorse possono modificare i comportamenti individuali. Questo non solo è noto, ma è anche auspicabile nella misura in cui vengono incoraggiati comportamenti virtuosi: se viene premiata la ricerca di qualità, probabilmente i ricercatori saranno indotti a migliorare il livello qualitativo del loro lavoro.

Una domanda che non ci si può esimere dal fare è, quindi, quali possibili conseguenze sui comportamenti ha l'adozione di un sistema di valutazione esclusivamente basato sugli indici bibliometrici. Nella misura in cui il parametro di valutazione può essere modificato dai comportamenti individuali, il suo uso genera, a sua volta, incentivi per comportamenti opportunistici da parte di singoli individui o gruppi di interesse<sup>5</sup>. Se si rifiuta, come argomentato in precedenza e come credo sia corretto fare, l'identità fra impatto e qualità della ricerca, la correlazione esistente fra le due grandezze rischia di affievolirsi.

Credo non sia difficile immaginare che l'uso esclusivo dell'Impact Factor potrebbe accrescere condotte strategiche da parte di individui, network di ricerca, riviste, in direzioni non necessariamente encomiabili. Penso, ad esempio, alla possibile spinta verso l'omologazione, specie dei giovani ricercatori; ad un aumento delle citazioni motivate da ragioni altre rispetto al riconoscimento di un debito intellettuale nei confronti dell'opera citata; ad un rischioso processo di selezione di temi ed autori da parte delle riviste.

In prospettiva, mi pare che le critiche di arbitrarietà e manipolabilità dei sistemi di valutazione diretta rischino di diventare pertinenti anche per gli indici bibliometrici. Con l'aggravante che per questi ultimi il sistema di referaggio anonimo ed il gran numero di soggetti coinvolti rende difficilmente controllabile l'insorgere di fenomeni degenerativi, già denunciati in varie sedi.

---

need to fix all the sloppiness - has led to our current misery. Moreover, the expectation that referees will rewrite papers, combined with sloppy submissions, makes refereeing extraordinarily unpleasant. We - the efficiency-obsessed academic discipline - have the least efficient publication process. The system is broken." Editor's Announcement : No Revisions Option.

<sup>4</sup> Un tema importante, non affrontato in questa nota, è che, soprattutto per alcune discipline, l'idea incorporata in un prodotto di ricerca non viene colta dagli indici bibliometrici, perché impatta su realtà diverse dalla comunità scientifica di riferimento (penso, ad esempio, ai brevetti, le forme organizzative, i prodotti artistici, ecc.).

<sup>5</sup> Si tratta di un problema assolutamente generale, che è stato ampiamente riconosciuto ed affrontato in numerosi campi di applicazione dei processi di valutazione. Per un singolare esempio, si legga l'articolo di Marc Santorra, New York Times, 11 gennaio 2005. Si legga anche Frey (2009).



5.

La valutazione della ricerca è utile e necessaria. I meccanismi di distribuzione delle risorse e di progressione delle carriere devono premiare la ricerca di qualità. Un impianto ben congeniato è equo ex post ed efficiente ex ante, nella misura in cui incentiva comportamenti virtuosi. Condizioni irrinunciabili di un buon sistema di valutazione della ricerca sono la sua certezza, la stabilità nel tempo e la non manipolabilità.

In questa prospettiva, gli indici bibliometrici sono un valido strumento di misurazione di alcuni aspetti della ricerca scientifica. In quanto semplici, rapidamente accessibili e poco costosi, essi costituiscono un utile indicatore nei processi di valutazione della ricerca. Tuttavia, occorre farne un uso consapevole. Le riflessioni svolte in questa breve nota evidenziano numerose criticità. Provo a riassumere quelle che mi paiono più rilevanti.

- a. Uno dei principali vantaggi del poter valutare i prodotti di ricerca associando loro un unico indice numerico è la semplicità. In particolare, il confronto degli indici consente un'immediata comparazione fra riviste, articoli, individui. Tuttavia, occorre essere consapevoli che l'obiettività di questo tipo di comparazione è discutibile. Una delle ragioni è che la contropartita della semplicità è costituita una perdita di informazioni che può essere rilevante al fine del giudizio comparativo. Inoltre, una tale pratica discrimina fra aree disciplinari e, all'interno di esse, specifici approcci e filoni di ricerca.
- b. Non è (statisticamente) lecito inferire la qualità di un articolo dalla qualità della rivista sulla quale è pubblicato. Se si vuole valutare il fattore d'impatto di un articolo/autore, è inappropriato utilizzare l'Impact Factor della rivista. Attenzione deve essere posta anche all'attendibilità dei dati utilizzati nel costruire misure individuali di impatto.
- c. E' sconsigliabile valutare la ricerca facendo esclusivo ricorso agli indici bibliometrici. Per almeno due motivi. Il primo è che gli indici bibliometrici non misurano, per stessa ammissione di chi li elabora, la qualità della ricerca (spesso questo fatto viene colpevolmente trascurato). Il secondo è che, sebbene i risultati ottenuti sulla base questi indici siano correlati positivamente con quelli prodotti da diversi metodi di valutazione (primo fra tutti i processi di peer review), il loro utilizzo esclusivo potrebbe incentivare comportamenti strategici non necessariamente virtuosi (su quali siano i comportamenti da incentivare nella comunità scientifica occorrerebbe una più ampia discussione) che, nel lungo periodo, rendono più labile il nesso fra qualità ed impatto. In questo senso, occorre essere ben consapevoli che le critiche di arbitrarietà e manipolabilità rivolte ai sistemi di peer review rischiano di rivelarsi pertinenti anche per gli indici bibliometrici.

Alcuni autori, partendo dai limiti delle misure bibliometriche, hanno sviluppato indici alternativi (si veda, ad esempio, la proposta di Marcuzzi e Zacchia, 2007). Certamente si tratta di strumenti di valutazione utili, che danno un importante contributo al dibattito in corso. Tuttavia, non nascondo un certo scetticismo nei confronti di queste alternative, se si pretende che esse costituiscano il criterio unico di giudizio<sup>6</sup>: è assai plausibile che l'uso esclusivo di un indice per valutare la ricerca metta in moto meccanismi strategici e comprometta la proprietà di non manipolabilità che un metro di misura dovrebbe rispettare, esattamente come discusso in precedenza in relazione all'Impact Factor.

In sintesi, la ricerca è un qualcosa di complesso. Cercare di ridurne l'essenza ad un unico indice numerico, sulla base della sua semplicità e supposta oggettività, potrebbe rivelarsi sbagliato, nella misura in cui ciò causa perdita di informazione rilevante. Potrebbe essere anche pericoloso, se induce comportamenti opportunistici volti alla manipolazione dei risultati.

---

<sup>6</sup> In questo senso va la proposta di Frey e Rost (2008).

### **Riferimenti Bibliografici**

Adler Robert, Ewing John (Chair), Taylor Peter (2008), Citation Statistics, report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS), 6/11/2008.

Artoni Roberto (2007), Valutazione della ricerca e pluralismo in Economia Politica, Rivista Italiana degli Economisti, XII, n. 2.

Marcuzzo Maria C. e Zacchia Giulia (2007), L'Econlit e gli strumenti per la valutazione della ricerca economica in Italia, Rivista Italiana degli Economisti, XII, n. 2.

Frey Bruno S. (2009), Economists in the PITS?, Institute for Empirical Research in Economics, University of Zurich, WP n. 406.

Frey Bruno S. e Katja Rost (2008), Do Rankings Reflect Research Quality?, Institute for Empirical Research in Economics, University of Zurich, WP n. 390.

Seglen Per O (1997), Why the impact factor of journals should not be used for evaluating research, BMJ; 314:497

Figura 1

