

Apples and oranges?

Spunti per una discussione sulla valutazione della ricerca nelle scienze umane e sociali

Antonio Banfi

Università degli Studi di Bergamo

Università degli Studi di Milano

email: antonio.banfi@unibg.it

Abstract: In this paper we consider the usefulness of bibliometric indicators for the evaluation of human and social sciences, with a special focus on the Italian exercise on evaluation. We analyze the configuration of human and social sciences as a whole, and we consider the disciplinary peculiarities of the HSS. We think that, while bibliometric analysis is theoretically appropriate for the HSS, the use of such indicators for many "soft" research fields often generates outcomes that are not statistically relevant.

Moreover, the cons (distortions, abuses, and so on) originated by the use of bibliometric indicators appear as more relevant than the pros, especially considering the limited funding needed for the HSS fields. A possible solution lies in the use of a more complex mix of indicators.

Keywords: scientometrics, Italian research assessment exercise, bibliometrics, human and social sciences, ANVUR, VQR.

In questi mesi ha preso avvio in Italia un complesso esercizio di valutazione della qualità della ricerca per il periodo 2004-2010 (VQR), sotto l'egida dell'Agenzia nazionale per la valutazione dell'università e della ricerca (ANVUR). L'esercizio fornisce un'occasione per riflettere sui problemi della valutazione della ricerca nelle scienze umane e sociali (d'ora in poi: HSS), in particolare per quanto riguarda l'impiego di strumenti di analisi bibliometrica in queste aree disciplinari. Una riflessione tanto più necessaria in quanto l'esercizio è partito in un contesto di quasi completa disinformazione e inesperienza circa i temi della valutazione, sia da parte dei soggetti valutati che da parte di buona parte dei valutatori.¹

1. Scienze umane e sociali: una distinzione utile?

In primo luogo vorrei affrontare la questione della definizione di quelle aree scientifiche che tradizionalmente o sono sottratte all'analisi bibliometrica per alcune caratteristiche ritenute loro proprie, oppure possono essere sottoposte a questo tipo di analisi solo con notevoli difficoltà.

E' uso parlare, per ragioni di brevità, di "scienze umane", anche se è più corretta la dizione "scienze umane e sociali". In alcune occasioni capita di sentir parlare di *soft* o *human sciences*, contrapposte alle cosiddette scienze dure (*hard sciences*).²

Stando all'attuale configurazione delle aree disciplinari nell'ordinamento italiano, da un canto vi sono le aree 1-9 CUN (matematica, fisica, chimica, scienze della terra, scienze biologiche, scienze mediche, scienze agrarie e veterinarie, ingegneria civile e

* Saggio predisposto nell'ambito del Gruppo di lavoro ASTRID su: "Il sistema universitario".

¹ Cfr. GALIMBERTI 2012.

² Cfr. MOED 2010, p. 147 ss.

architettura, ingegneria industriale e dell'informazione), dall'altro le aree 10-14 (scienze dell'antichità, filologico letterarie e storico artistiche, scienze storiche, filosofiche, pedagogiche e psicologiche, scienze giuridiche, scienze economiche e statistiche, scienze politiche e sociali).

A ben vedere l'opposizione fra discipline *hard* e *soft* costituisce una semplificazione di scarsa utilità quando si considerino le differenze quanto a metodologia e *output* della ricerca all'interno delle diverse aree.

Tanto per fare qualche esempio, all'interno del “mare” delle scienze dure si trovano alcuni “isolotti” scientifici che non senza qualche buona ragione si potrebbe essere tentati di ascrivere alle HSS: didattica e storia della fisica (FIS 08), storia della medicina (MED 02), storia dell'architettura (ICAR 18) e forse anche restauro (ICAR 19), almeno per la stretta parentela di questo settore con la storia delle tecniche artistiche.

Si tratta infatti di discipline storiche, per quanto strettamente connesse al mondo delle *hard sciences*. Difficile ipotizzare che metodologia e esiti della ricerca siano, per queste discipline, del tutto omogenei a quanto accade per gli altri settori disciplinari della medesima area.

Lo stesso discorso si potrebbe fare per le aree del gruppo delle HSS. Vi sono, infatti, settori disciplinari che condividono pratiche assai simili a quelle proprie delle scienze dure: ad esempio la produzione prevalente di articoli su rivista, l'uso comune della lingua inglese, una limitata produzione di studi monografici.

Basti pensare, per questi aspetti, al complesso delle scienze economiche e statistiche (escluse la storia del pensiero economico e la storia economica, che in questo contesto si presentano come una sorta di *enclave* dotata di pratiche scientifiche autonome)³, che non a caso condividono già da tempo un approccio alla valutazione di tipo bibliometrico; oppure si pensi alle scienze linguistiche, glottologiche, alla logica e filosofia della scienza (M-FIL 02), tradizionalmente imparentata con la logica matematica (MAT 01), alle scienze psicologiche (che condividono alcuni aspetti con le discipline mediche) e a quelle pedagogiche. Ma qualcosa del genere potrebbe forse dirsi anche per le scienze archeologiche (i risultati di uno scavo assai spesso non sono pubblicati in una monografia in lingua nazionale), ed anche per archivistica, bibliografia e biblioteconomia. Anche una parte rilevante dell'area 14 (scienze politiche e sociali) potrebbe essere aggiunta a questo computo.

Del resto, anche le *hard sciences* non condividono uniformi stili citazionali: *time is not the only factor causing systematic differences in samples of publication and citation data. Different disciplines have innate, cultural differences in the way in which they use the literature, in terms of article length, frequency and citation structures. In crude terms, biomedical researchers tend to produce more, shorter papers where methodology and prior knowledge are extensively codified in a dense array of citations. Physical scientists and engineers produce less frequent but longer papers, with more detailed content and fewer cross-references. These characteristics, not relative quality, affect typical citation rates.*⁴

³ Vedi in proposito BACCINI 2012.1.

⁴ UNIVERSITIES UK, *Bibliometrics*, p. 8.

2. Uniformità o varietà?

Tutto ciò indica che il complesso delle discipline scientifiche presenta una realtà molto più complessa di quanto normalmente si immagina; se però si scende ancor più in profondità, tenendo presente l'obiettivo della valutazione e analizzando le pratiche disciplinari, emerge in modo ancora più evidente come all'interno di una stessa area vi siano rilevanti differenze fra settori: *a practical challenge in evaluation-oriented scientometrics is to cope with the consequences of this diversity—namely, the heterogeneity of areas— and to find appropriate reference sets for indicators to avoid mixing apples and oranges.*⁵

Ipotizziamo che si condivida il postulato secondo il quale nessuna area disciplinare è intrinsecamente non misurabile dal punto di vista bibliometrico.

E' un assunto del tutto corretto: la bibliometria è uno strumento, una metrica della scienza, pertanto qualsiasi disciplina che produca un *output* può essere in qualche modo misurata. Lasciamo per un momento da canto tutti gli aspetti problematici relativi alla bibliometria nelle HSS, e partiamo dunque da questo presupposto.

E' uso misurare i dati citazionali di singoli prodotti della ricerca (di norma articoli) o di intere riviste per determinarne, attraverso la costruzione di vari indicatori, l'impatto sulla comunità scientifica. A sua volta un elevato impatto può, pur con molte cautele, essere considerato come indice di qualità di un prodotto, anche se non è vero il contrario.

Prendiamo ora come campione un'area disciplinare alla quale si tende generalmente ad attribuire un buon grado di omogeneità: le scienze giuridiche.

E' evidente a chiunque che un tempestivo *paper* relativo alla riforma dell'articolo 18 dello Statuto dei lavoratori, o uno relativo alla sentenza della Corte Europea dei Diritti dell'Uomo in materia di immigrazione clandestina, avranno una ragionevole *chance* di attrarre un ampio numero di citazioni entro un tempo non troppo lungo, in quanto intervengono su di una materia attuale e controversa. Il giuslavorista, penalista o costituzionalista potrà agevolmente migliorare i propri indici citazionali, anche a vantaggio delle proprie strutture di appartenenza. Un fenomeno ulteriormente amplificato dalla diffusione di riviste scientifiche *on-line* che intervengono tempestivamente su temi di attualità.

La situazione muta drasticamente se pensiamo a un saggio in materia di *cognitio extra ordinem* di età imperiale (diritto romano e diritti dell'antichità), a uno su Accursio (storia del diritto medievale e moderno), o a uno scritto sul pensiero di Pufendorf (filosofia del diritto).

All'interno della stessa area, pur in presenza di una relativa omogeneità di metodo, i singoli settori si caratterizzano per esiti diversi e per diverse pratiche citazionali: il romanista, per scegliere uno fra i tre personaggi della nostra rappresentazione, citerà forse una risalente letteratura tedesca che ritiene molto autorevole, cosa che il giuslavorista non riterrà minimamente di fare, se non - come talora capita - per scopi meramente esornativi.

⁵ ZITT 2005, p. 40.

Il giurista positivo, infatti, intervenendo su temi connessi all'attualità e al diritto vivente, farà uso di stili citazionali adatti allo scopo. D'altra parte, mentre il giuslavorista otterrà numerose citazioni nella letteratura scientifica in tempi brevi e magari sarà perfino ripreso dalla stampa, contribuendo così alla "terza missione" dell'università, il nostro romanista sarà citato -forse- in tempi molto lunghi, da qualche altro studioso interessato all'epoca e agli istituti da lui analizzati. Da un canto vi è chi interviene sull'attualità e su "urgenze" diffusamente percepite, dall'altro chi lavora su una ricerca semplicemente *curiosity driven*. Sono comparabili questi soggetti? Credo di no.

A questo punto si rende opportuna qualche rapida osservazione circa l'analisi bibliometrica. Perché si fa ricorso a questi strumenti? Essi servono in primo luogo a rendere più rapido e meno costoso il processo di valutazione (la revisione dei pari richiede infatti tempi più lunghi ed è più onerosa, visto l'impegno che richiede ai revisori).

L'analisi bibliometrica è di importanza rilevante non tanto per la valutazione dei singoli, quanto per orientare le politiche pubbliche. Specie in un contesto di risorse scarse, il finanziamento alla ricerca dev'essere orientato nel modo migliore, perché "frutti" il più possibile, riducendo la dispersione dei fondi.⁶ Come spiega Alberto Baccini, "la valutazione avviene invariabilmente in due forme diverse: (i) la revisione diretta da parte dei pari, come nel caso in cui si richieda a panel di esperti di valutare la qualità di prodotti o di progetti di ricerca; (ii) la revisione indiretta dei pari, quando si valuta la qualità di un prodotto di ricerca raccogliendo, tramite l'osservazione sistematica delle pratiche adottate da una data comunità scientifica, i segnali della valutazione di qualità effettuata direttamente dai pari per fini diversi dalla valutazione. È il caso, per esempio, dell'uso di misure indirette di qualità basate sulle citazioni ricevute da un certo contributo scientifico".⁷

In effetti, come osserva anche Donald Gillies,⁸ l'analisi citazionale è, almeno in alcuni ambiti scientifici, una misura indiretta dell'opinione dei pari, rilevabile attraverso il ricorso a opportune basi dati: da questo punto di vista, la distinzione fra oggettività propria dell'indagine bibliometrica e soggettività tipica della *peer review*, è per certi versi illusoria, poiché la somma delle citazioni, almeno per certe discipline può essere considerata come rappresentativa dell'opinione dei pari; con il rischio di un ulteriore cortocircuito: *instead of subjective peer review one has the subjective interpretation of a citation's meaning. Those who promote exclusive reliance on citation-based metrics implicitly assume that each citation means the same thing about the cited research its "impact". This is an assumption that is unproven and quite likely incorrect.*⁹

In ogni caso, vale la pena di ribadirlo, i sistemi di analisi bibliometrica, per quanto nati per facilitare le scelte dei responsabili degli acquisti delle grandi biblioteche, non sono stati sviluppati per le HSS, bensì per le scienze dure, quelle per cui la

⁶ Cfr. GEUNA, MARTIN 2003.

⁷ Cfr. BACCINI 2011.

⁸ GILLIES 2008, p. 55 ss.

⁹ ADLER, EWING, TAYLOR 2008, p. 4; JENSEN, ROUQUIER, CROISSANT 2008.

corretta allocazione delle risorse da parte dello Stato (o dei finanziatori in genere) assume maggior rilievo, almeno in termini di trasferimento tecnologico.

Nel momento in cui si decide di applicare in modo più o meno meccanico all'ambito delle HSS sistemi di analisi già sperimentati da tempo in contesti diversi, occorre pertanto esercitare cautela ed essere pienamente consapevoli delle pratiche scientifiche che caratterizzano i singoli settori disciplinari sottoposti a valutazione.

3. Concentrare la ricerca per incrementarne l'output?

L'eventuale ricorso a indici citazionali come strumento di valutazione nelle aree umanistiche comporta la necessità di un mutamento nello stile di ricerca di questi settori (favorendo la pubblicazione di articoli su riviste indicizzate a scapito delle monografie), oppure l'applicazione di strumenti di analisi citazionale anche ai lavori monografici, cosa tutt'altro che semplice.

A questo proposito, va rilevato come l'analisi citazionale tende a favorire non solo la ricerca *mainstream* (intesa come ricerca prevalente condotta con metodi condivisi e consolidati), ma anche e soprattutto i filoni "alla moda"¹⁰ e così pure il cosiddetto *groupthinking*.¹¹ Tralasciando per il momento le giustificate critiche rivolte per questo e per altri aspetti a tali strumenti di valutazione, è comprensibile che, ad una prima impressione, il fenomeno della concentrazione della ricerca su determinati filoni possa essere valutato positivamente: se la ricerca deve essere *stricto sensu* produttiva, niente di male nel fatto che i ricercatori intervengano sui temi più "caldi" del momento, quelli che evidentemente sollecitano la maggior concentrazione e il maggior sforzo da parte della comunità scientifica.

Si tratta però di una visione schematica del progresso scientifico, che sottovaluta l'importanza della ricerca *curiosity driven* (assai rilevante anche nelle scienze dure) e che riflette aspetti propri di alcune specifiche aree disciplinari, o per meglio dire di alcuni settori disciplinari.

E', insomma, una visione modellata sulla cosiddetta ricerca *agenda driven*, che a differenza di quella *curiosity driven* produce i suoi effetti in tempi solitamente più rapidi, a scapito però dell'innovatività.

E' evidente che possono eventualmente esservi dei benefici nel concentrare gli sforzi dei ricercatori, ad esempio nell'analisi di una nuova proteina che promette ottimi risultati per la cura di certe malattie; così pure è ovvio che molti studiosi di diritto costituzionale potranno opportunamente decidere di concentrare le loro energie sull'esame della questione dei diritti delle persone, in una fase di trasformazione delle strutture sociali e familiari. Similmente molti economisti vorranno concentrarsi sul problema delle forme dell'intervento statale in economia, nell'attuale contingenza storica.

¹⁰ Con ciò si intende quei filoni di ricerca associati a elevato numero di pubblicazioni (e citazioni) in un lasso di tempo relativamente breve.

¹¹ Ossia un pensiero unanimistico e coesivo di gruppo, un fenomeno tutt'altro che estraneo al mondo accademico

Tuttavia, vi sono numerose discipline nelle quali è difficile che si producano fenomeni di significativa concentrazione della ricerca intorno a filoni oggetto di un'attenzione condivisa da parte di settori rilevanti della comunità dei ricercatori. Penso, fra le altre, alle discipline storiche (con l'eccezione, forse, di quelle contemporaneistiche, per le quali un'eventuale contrapposizione di carattere ideologico intorno a determinati oggetti di studio potrebbe agire da volano per la moltiplicazione delle citazioni) e a buona parte di quelle filologiche e letterarie.

In questi ambiti, mentre le metodologie adottate dai ricercatori sono generalmente costanti e condivise, oppure evolvono solo nel lunghissimo periodo, la ricerca procede in modo per così dire atomizzato, seguendo gli interessi dei singoli ricercatori, il che ha delle conseguenze non irrilevanti circa i tempi (più lunghi) e la probabilità (più scarsa) di ottenere citazioni.

In un contesto di questo genere, la misurazione dei dati citazionali rischia non solo di essere distorsiva, ma di produrre dei risultati inutilizzabili. L'analisi citazionale richiede, per essere attendibile e fornire informazioni di una qualche utilità, che il numero delle citazioni potenziali di un determinato prodotto della ricerca sia almeno relativamente elevato. Insomma, numeri troppo piccoli rendono insignificante l'analisi citazionale; il confronto fra un articolo oggetto di cinquanta citazioni e uno oggetto di dieci citazioni può forse dire qualcosa circa il diverso impatto dei due prodotti sulla comunità scientifica. Lo stesso non può dirsi per numeri e differenze troppo piccole: il confronto fra due articoli citati rispettivamente una e tre volte, non è significativo e dice molto poco sull'impatto di questi prodotti. Si è pensato di ovviare alle differenze fra gli stili citazionali dei diversi settori ricorrendo a normalizzazioni (si sa, per esempio, che i fisici si citano fra loro molto più dei matematici).

Eppure, anche se si effettua una normalizzazione per settore disciplinare, può ben darsi il fenomeno per cui una singola citazione, in un contesto complessivo di scarse occorrenze di citazioni, è sufficiente per determinare l'attribuzione del prodotto scientifico ad una diversa classe di qualità: capiterà così che, volendo suddividere per classi i prodotti della ricerca, un prodotto definito come eccellente si segnali per 3 citazioni, uno buono per 2, uno mediocre per 1 sola citazione, e uno limitato perché mai citato.

Come ha scritto J. Adams, *citation growth characteristics vary between fields, and larger fields tend to have higher citation rates*,¹² ed infatti il fenomeno si produce anche nelle "scienze dure".¹³ Tuttavia, poiché la ridotta dimensione di un settore comporta una riduzione nel numero di citazioni, tale evenienza si potrebbe presentare più frequentemente nell'ambito delle HSS; inoltre questo fenomeno nelle HSS è ulteriormente favorito dal fatto che la ricerca in questi ambiti disciplinari è spesso condotta individualmente al di fuori di filoni d'indagine condivisi da gruppi di ricercatori di dimensioni significative.

¹² Cfr. ADAMS 2007; anche RADICCHI, FORTUNATO, CASTELLANO 2008.

¹³ Cfr. PRATELLI, BACCINI, BARABESI, MARCHESELLI 2011.

Che in contesti di questo tipo i dati citazionali consentano di apprezzare l'effettivo impatto dei diversi prodotti della ricerca resta pertanto da dimostrare: infatti, la loro significatività statistica è assai bassa.

Ad esempio, uno studioso di papirologia giuridica, disciplina di notevole complessità tecnica e fortemente internazionalizzata, che ha portato notevoli avanzamenti nella storia del diritto antico, ha davanti a sé un numero di citazioni *potenziali* per i suoi scritti estremamente ridotto; egli potrà raggiungere il massimo delle proprie citazioni possibili con un punteggio assai limitato, a prescindere dalla qualità intrinseca dei propri lavori: gli studiosi di papirologia giuridica sono infatti pochi e orientano le loro indagini nelle direzioni più diverse sulla base di preferenze e inclinazioni soggettive. E' attendibile, utile e opportuna un'analisi citazionale per ambiti di questo tipo?

Ha senso adottare per un insieme eterogeneo di ricercatori il medesimo criterio di misura sulla base dell'appartenenza alla medesima area oppure al complesso delle HSS? E soprattutto, siamo certi che una scelta del genere non finisca per danneggiare una parte rilevante delle discipline umanistiche?

Le pratiche e le abitudini sopra descritte non sono infatti di per sé segno di arretratezza o provincialismo, poiché caratterizzano pressoché ovunque nel mondo il modo di procedere di alcune discipline umanistiche.

Del resto, l'idea di orientare, attraverso gli strumenti di valutazione, gli sforzi dei ricercatori prevalentemente in certe direzioni, causando una sorta di "concentrazione", può essere criticata come per più aspetti sbagliata e perniciosa con riguardo a tutte le discipline, in quanto simili pratiche rischiano di ridurre il pluralismo e isterilire la ricerca; rischio che è per forza di cose ancora maggiore per quelle discipline nelle quali la ricerca è per antonomasia *curiosity driven*, intrinsecamente atomizzata e per sua natura priva di filoni d'interesse *mainstream* ben definiti.

In tal senso, perfino la proposta – apparentemente ragionevole – di costruire basi dati specificamente tarate su singole aree disciplinari, pone più problemi di quanti ne risolve. Basti pensare all'area giuridica (area CUN 12): non si può pensare di assimilare completamente le abitudini di ricerca, l'*output* e le abitudini citazionali dei 21 settori disciplinari giuridici che compongono l'area. Quanto hanno in comune, per questo profilo, diritto amministrativo, storia del diritto medievale e moderno, diritto della navigazione, diritto comparato, diritto romano, diritto internazionale e filosofia del diritto? Ragionamenti analoghi potrebbero probabilmente essere condotti per numerosi settori delle altre aree "umanistiche".

Come hanno giustamente scritto E. Giménez-Toledo e A. Román-Román, *we are used to looking at humanities and social sciences as a whole, when in reality it is a group of disciplines, each one very distanced from the other as much for the aims of their study as for their specific cultures, their work materials, their methods of research, etc. In addition, if comparability is sought, it is necessary to work by thematic areas. It is also important to point out that the difficulty with some of the mentioned approaches,*

*citation analysis for instance, varies depending on what area of study we are talking about.*¹⁴

In verità, il ricorso a strumenti bibliometrici può finire per ridurre alcuni settori delle HSS a un ruolo meramente ancillare rispetto ad altre discipline il cui maggiore impatto citazionale corrisponde a ricadute socio-economiche più visibili e più facilmente apprezzabili nel breve termine.¹⁵ Non so davvero dire se ciò comporterà un beneficio in termini di effettiva qualità della ricerca, poiché essa si nutre anche di pluralismo e beneficia della varietà degli interessi dei singoli ricercatori. Quali conclusioni trarre da quanto detto fin ora?

1. le HSS non costituiscono un insieme omogeneo come talora si pretende.
2. vi sono alcune discipline per le quali l'analisi bibliometrica, per quanto teoricamente possibile, non è in grado di condurre a risultati significativi e dunque in qualche modo utilizzabili sotto il profilo della valutazione, a causa del basso numero di citazioni complessive e della loro conseguente scarsa significatività statistica.¹⁶
3. l'adozione di metriche citazionali rischia per alcuni settori delle HSS di risolversi nella marginalizzazione di determinate discipline, all'interno di aree più ampie. Marginalizzazione determinata non dallo scarso valore dei prodotti della ricerca, ma da peculiari costumi citazionali, struttura delle comunità di riferimento e pratiche scientifiche autonome, per quanto esse siano condivise anche a livello internazionale.

In verità, non è solo il diverso comportamento dei ricercatori in molte delle HSS e la complessità dell'insieme delle HSS a suggerire cautela nell'adozione di modelli di valutazione analoghi a quelli previsti per le *hard sciences*. Occorre anche valutare bene se i costi in senso lato, comprensivi dunque dei danni eventualmente apportati a questi campi di ricerca, siano proporzionati ai ricavi. Di questo si dirà qualcosa più avanti. Resta fermo, tuttavia, che non è pensabile che le HSS debbano sottrarsi alla valutazione e, grazie ad essa, a un miglioramento della qualità complessiva della ricerca. Ma quale valutazione?

4. Competere, internazionalizzarsi: è tutt'oro quel che luccica?

L'obiettivo dell'Agenzia italiana di valutazione, è di assicurare un rapido riallineamento della ricerca scientifica italiana ai migliori livelli internazionali. Sul punto occorre premettere qualche riflessione. Va infatti ricordato ancora una volta che il livello della ricerca scientifica, stando ai dati Scopus, è di notevole rilievo e anzi migliore di quanto ci si potrebbe attendere in considerazione dell'evoluzione della situazione economica nazionale e degli scarsi finanziamenti erogati.¹⁷ Questo

¹⁴ GIMÉNEZ-TOLEDO, ROMÁN-ROMÁN 2009, p. 202.

¹⁵ Va notato che definire che cosa si intenda per "impatto" (in senso non strettamente citazionale) non è un compito semplice, anche perché tale definizione potrebbe variare da disciplina a disciplina. Cfr. MARTIN 2011, p. 249.

¹⁶ Cfr. BACCINI 2012.2.

¹⁷ In proposito cfr. DE NICOLAIO 2012.

tipo di considerazioni non possono però essere automaticamente estese dall'ambito delle scienze dure a quello delle HSS, proprio per le notevoli difficoltà di misurarne esattamente la produzione: Scopus infatti misura prevalentemente la produzione delle "scienze dure".

Si può supporre che il livello qualitativo della produzione scientifica nelle HSS non si discosti di molto, in proporzione, da quello delle *hard sciences*, ma non ve ne è prova; il che rende ancora più opportuno sostenere l'adozione delle "migliori pratiche" da parte degli studiosi di queste aree, in modo da migliorarne ulteriormente la qualità.

Occorre però essere cauti e dedicarsi a una riflessione attenta, per evitare esiti indesiderati: ricette troppo semplici potrebbero in quest'ambito produrre pietanze immangiabili.

Una buona competizione fra ricercatori può essere senz'altro auspicabile. Non sono certo, però, che l'analisi citazionale possa favorirla, anzi. Come già si è detto, in molti settori e discipline delle HSS la granularità dei dati citazionali è grossolana. Ciò non solo non rende effettivamente apprezzabile l'impatto di un contributo in tempi ragionevoli, ma rischia di favorire i comportamenti opportunistici già verificatisi in alcune aree delle scienze dure. Vediamo perché.

E' ben noto che alcuni sistemi di valutazione rischiano di favorire effetti distorsivi; fra questi merita di essere ricordato in primo luogo l'effetto S. Matteo (o *Matthew effect*). Esso prende il nome da un versetto del Vangelo di Matteo (25. 29: "Perché a chiunque ha sarà dato e sarà nell'abbondanza; ma a chi non ha sarà tolto anche quello che ha") e descrive il fenomeno per cui l'analisi quantitativa può rafforzare la tendenza della scienza a organizzarsi secondo filoni di ricerca omogenei scoraggiando di fatto l'innovazione e lo sviluppo stesso della ricerca scientifica verso nuove vie inesplorate. Continuare a studiare su temi già noti e ampiamente discussi nella comunità scientifica, assicura infatti una buona "resa" bibliometrica, ma al contempo impoverisce la ricerca e la irrigidisce caratterizzandola per un comportamento inerziale: si prosegue, insomma, lungo la stessa strada, senza deviare da essa.

Gli effetti possono anche essere gravi, come risulta da studi condotti nell'ambito delle scienze dure. Si pensi alla chimica e alle scienze ambientali: secondo un ben documentato studio apparso su *Environmental Health*, *the results obtained [...] suggest a serious bias in publications on environmental chemicals. While regulatory agencies request information on poorly studied, potentially serious environmental hazards, the publications in scientific journals emphasize a relatively small, selective number of highly prominent chemicals, about which large numbers of articles were published already during the past century.*¹⁸

In questo contesto vale la pena di considerare anche la cosiddetta legge di Goodhart: nata in ambito economico, sancisce nelle sue più recenti formulazioni che *when a measure becomes a target, it ceases to be a good measure*. Applicata all'ambito scientometrico, la legge descrive il modo in cui, posti degli obiettivi, essi cessano di

¹⁸ Cfr. GRANDJEAN, ERIKSEN, ELLEGAARD, WALLIN 2011

essere una misura della qualità della ricerca, per divenire un obiettivo perseguito di per sé. Ciò, detto in altre parole, significa che determinati strumenti di valutazione vengono distorti nel momento in cui i soggetti che sono sottoposti alla valutazione adottano comportamenti opportunistici per ottenere un determinato punteggio a prescindere dall'effettiva qualità della ricerca.

Ad essa si ricollega la legge di Campbell: *the more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor*. E' noto, ad esempio, che nel Regno Unito, nel momento in cui si assunse come criterio di giudizio dell'impatto della produzione accademica il numero di *spin offs*, che avrebbero dovuto indicare un elevato trasferimento tecnologico, si moltiplicarono, favorite dalle stesse università, *spin offs* che non ammontavano a nulla di più di un marchio registrato e di una singola unità di personale (il fondatore).¹⁹

Douglas N. Arnold e Kristine K. Fowler, hanno scritto un ben noto *paper*, intorno alle manipolazioni dell'*impact factor* delle riviste, con considerazioni che possono essere facilmente estese all'ambito della valutazione individuale: *it is possible to vastly increase impact factor without increasing journal quality at all. The actions of a few interested individuals can make a huge difference, yet require considerable digging to reveal*.²⁰ Si sa che per questo motivo nell'attività di valutazione ci si trova davanti ad alcuni problemi non irrilevanti da risolvere; a partire dal fenomeno delle autocitazioni, la cui esclusione richiede un impegno consistente (traducibile in costi crescenti).²¹ Va detto che un serio meccanismo di *peer review* attuato al livello delle riviste scientifiche potrebbe limitare comportamenti di questo genere, ma rimane il problema dei coautoraggi falsi o pretestuosi, che però ha, almeno finora, molto maggior rilievo per le scienze dure che per le HSS, dove sono piuttosto rare le pubblicazioni con più autori (anche se vi sono settori che fanno eccezione, a partire dalle scienze economiche e statistiche).

Resta invece forte il rischio che si sviluppino comportamenti abusivi, con la formazione di "cerchie" di soggetti pronti a citarsi l'un l'altro, e eventualmente a escludere dalle proprie citazioni i gruppi rivali. Fenomeni di questo genere sono ben noti e hanno favorito l'emergere di posizioni assai critiche anche all'interno delle "scienze dure", suscitando talora reazioni esasperate.²²

Tanto più che alcuni sistemi di misurazione dell'impatto, specie quelli fondati su Google Scholar, hanno già dimostrato di essere manipolabili, tanto da consentire addirittura la costruzione di scienziati inesistenti. Come ha osservato Labbé a

¹⁹ MARTIN 2011, p. 249.

²⁰ ARNOLD, FOWLER 2011, p. 6.

²¹ Sulle autocitazioni cfr. fra gli altri CASSELLA, BOZZARELLI 2011.

²² ERNST 2010 p. 90: *and as an ultimate plea, the personal wish of the author remains to send all bibliometrics and its diligent servants to the darkest omnivoric black hole that is known in the entire universe, in order to liberate academia forever from this pestilence. – And there is indeed an alternative: Very simply, start reading papers instead of merely rating them by counting citations!* Cfr. anche MOLINIÉ, BODENHAUSEN 2010, pp. 78-89; sulla questione vedi anche DE NICOLAIO 2011.

proposito dello scienziato fasullo Ike Antkare, al quale si è riusciti a far ottenere un h-index pari a 84 fondato su *papers* fasulli, *this distortion could have been easily achieved using names of real people, thus fostering or rather discrediting them.*²³

Vorrei sottolineare come la questione sia particolarmente seria nel campo delle HSS: in un ambito accademico con un numero molto ampio di ricercatori e di conseguenza con un vasto numero di prodotti della ricerca, vi saranno dati citazionali consistenti, il che non potrà annullare le distorsioni, ma almeno contribuirà a ridurre il peso complessivo. Al contrario, in alcuni settori delle HSS, dove vi sono pochi ricercatori e una produzione relativamente ridotta, la possibilità di alterare i dati citazionali attraverso accordi di consorterie *anche di ridottissime dimensioni* diviene molto concreta, e anzi va considerata come una quasi-cerchezza: insomma, “citiamoci a vicenda e scaliamo le vette delle classifiche bibliometriche”. Va ricordato che la manipolazione dei dati citazionali e di impatto è ormai un problema di primaria importanza nell’analisi bibliometrica.²⁴

Ancora, di recente è stata pubblicata una ricerca secondo la quale *a number of business and economic journals systematically manipulate their impact factors and try to artificially boost their academic reputation,*²⁵ tanto che si è potuta compilare una classifica dei *rogue journals*.²⁶

Table S12. Journals identified as coercers by survey respondents. Number of coercive observations represents the number of times a journal was identified by independent survey respondents as requesting self citations that (i) give no indication that the manuscript was lacking in attribution, (ii) make no suggestion as to specific articles, authors, or a body of work requiring review, and (iii) only guide authors to add citations from the editor’s journal. Ties are listed alphabetically.

Journal Title	Number of Coercive Observations
<i>Journal of Business Research</i>	49
<i>Journal of Retailing</i>	43
<i>Marketing Science</i>	29
<i>Journal of Banking and Finance</i>	24
<i>Information and Management</i>	19
<i>Applied Economics</i>	18
<i>Academy of Management Journal</i>	14
<i>Group and Organization Management</i>	13
<i>Journal of Consumer Psychology</i>	9
<i>Psychology and Marketing</i>	8
<i>Journal of Accounting and Public Policy</i>	7
<i>Journal of Financial Economics</i>	7
<i>Academy of Management Review</i>	6

Se queste sono le distorsioni che la comunità scientifica può riconoscere per settori significativamente numerosi, quali esiti dobbiamo attenderci per alcune fra le HSS, caratterizzate da una scarsa popolazione di ricercatori?

²³ LABBÉ 2010, p. 50.

²⁴ Cfr. VAN RAAN 2006.

²⁵ Cfr. STORBECK 2012.

²⁶ Cfr. WILHITE, FONG 2012.

I comportamenti opportunistici sono infatti favoriti dalle ridotte dimensioni, in termini di ricercatori e dunque di *output* complessivo, del gruppo disciplinare preso in considerazione.

Bisogna dunque chiedersi se l'adozione di strumenti di valutazione che facciano ricorso all'analisi citazionale per settori disciplinari di dimensioni ridotte non possa costituire un potente incentivo all'adozione di pratiche opportunistiche che, per la dimensione del gruppo di ricercatori oggetto di valutazione, potrebbero essere in grado di orientare in modo efficace gli esiti della valutazione stessa, limitando al contempo la possibilità di individuare e porre efficacemente rimedio al fenomeno.

E' un fenomeno noto, che ha spinto gli studiosi di bibliometria a elaborare altri indici, per ovviare agli inconvenienti dell'h-index.²⁷ Il che però ci porterebbe forse in un labirinto bibliometrico di cui si rischierebbe di perdere il senso e l'utilità. Si tratta infatti di discipline che per lo più non necessitano di finanziamenti per la ricerca di dimensioni così rilevanti da giustificare di per sé il ricorso a metriche troppo sofisticate e alla costruzione di costose basi dati. Se si ammette che nelle HSS, come afferma l'Accademia Reale d'Olanda, *there can be no question of automatic calculations*,²⁸ resta da capire quale sia l'effettivo valore delle informazioni bibliometriche e se il costo della loro estrazione sia pienamente giustificato, al di là di meri esercizi di *testing*.

Un'altra questione che riguarda i ricercatori delle HSS, nasce dalla consolidata pratica degli studiosi, specie più giovani, di dimostrare la propria diligenza attraverso una minuziosa e il più possibile accurata rassegna della dottrina precedente. Questa abitudine citazionale potrebbe costituire un ostacolo nella elaborazione degli indici citazionali per le HSS. Però, su questo punto, l'obiezione è molto meno forte di quella concernente le auto-citazioni, e lo stesso vale per la questione delle citazioni "critiche", che sono molto più frequenti nelle HSS che nelle scienze dure. L'uso di infarcire le prime pagine di articoli e monografie di note di dimensioni abnormi contenenti la diligente sintesi della bibliografia in materia non giova in alcun modo alla ricerca e andrebbe scoraggiato. Quanto alle citazioni critiche, è facile sostenere che un'opera fortemente criticata e discussa ha comunque esercitato un impatto di un qualche rilievo sulla comunità accademica. Del resto, la stessa comunità scientifica, se consapevole dell'applicazione di determinati strumenti, potrebbe opportunamente rinunciare alla citazione critica di lavori *impresentabili*, per riservare la propria attenzione a opere *controverse*. A questo proposito il *Research Report* di Universities UK, *The use of bibliometrics to measure research quality in UK higher education institutions*,²⁹ ricorda che *there is frequent concern that some papers accumulate significant citation counts 'because they are wrong'. There is little evidence of this. The bulk of ill-conceived work that does pass editorial scrutiny and reach publication probably remains uncited because it is also trivial. A wellknown example of frequently cited 'wrong' work (Fleischman and Pons on cold fusion) still had a significant and non-trivial impact on work in its field.*

²⁷ VAN RAAN 2006, p. 499 ss.

²⁸ QUALITY INDICATORS 2011, p. 12.

²⁹ UNIVERSITIES UK, *Bibliometrics*, p. 24.

Un ultimo punto che merita attenzione è il problema delle ricerche interdisciplinari³⁰; esse sono proprie di tutte le aree scientifiche, ma sono particolarmente radicate nelle HSS. Per di più si tratta di un tipo di indagine specialmente rilevante per il progresso complessivo dell'attività di ricerca, poiché l'interdisciplinarietà può agire da potente "fertilizzante" per lo sviluppo di analisi complesse: *multidisciplinarity is a crucial phenomenon in the evolution of science, with many interpretations along with its variants, interdisciplinarity and transdisciplinarity. It can be seen as both the counterpart of diversity—maintaining connections between established domains—and the source of further diversity—creating new buds by recombination.*³¹ Eppure, l'analisi bibliometrica per questo aspetto rimane per più di un verso incapace di fornire un quadro completo della situazione: *scientometrics offers some powerful means to help with the understanding of multidisciplinarity, a quite difficult subject touching on the very structure and dynamics of science. Some modest achievements have already been shown by scientometricians in various contexts, but, clearly, developing measures adapted to emerging research questions is still a challenge for the discipline.*³²

Torniamo dunque al punto: è bene che la ricerca italiana nelle HSS faccia ogni sforzo per progredire e per migliorare. Occorre però porsi ancora qualche domanda. Che cosa vogliamo ottenere? E a cosa serve l'analisi bibliometrica? Proviamo rispondere sottolineando alcuni punti.

5. Quale futuro per la valutazione delle HSS in Italia?

Vorrei tentare di tratteggiare, senza alcuna pretesa di sistematicità e completezza, alcuni possibili percorsi.

Valutare la mera produttività di un ricercatore delle HSS non è cosa difficile. Basta definire chiaramente il campo dei prodotti valutabili come aventi valore scientifico e contarli. Un compito che si potrebbe esaurire in pochi giorni senza gran dispendio di risorse. Altra cosa è favorire le migliori pratiche e stimolare la ricerca nelle HSS affinché essa migliori il suo *status* in ambito internazionale. Per questo aspetto si possono avanzare alcuni suggerimenti che vanno però accompagnati da altrettante avvertenze.

Come si è cercato di dimostrare, il ricorso all'analisi citazionale e ad altre metriche dell'impatto è, per molte HSS, del tutto discutibile: molti prodotti della ricerca hanno peraltro un impatto lento e spalmato sul lungo periodo. Orientare i ricercatori verso prodotti di rapido e facile impatto trasformandoli in cacciatori di citazioni

³⁰ Uso il termine in senso lato senza entrare nelle fini distinzioni fra ricerca interdisciplinare, multidisciplinare e transdisciplinare.

³¹ ZITT 2005, p. 44.

³² *Ibid.* p. 46

comporterebbe per questi settori un danno maggiore di quello già visibile in alcune aree scientifiche.

Più efficaci potrebbero essere azioni che favoriscano la redazione di una quota della produzione scientifica in lingua inglese, purché ben calibrate e adottate con giudizio. E' noto a tutti che l'inglese svolge in questa fase storica un ruolo simile a quello che ebbe a suo tempo il latino: adottare la "lingua franca" almeno per una certa parte delle pubblicazioni può contribuire in modo efficace ad aumentare la diffusione dei prodotti della ricerca italiani e accrescere il loro impatto sulla ricerca a livello globale.

Inutile dire che imporre sistemi che valutino automaticamente come di miglior qualità un prodotto della ricerca in quanto scritto in lingua diversa da quella nazionale sarebbe una sciocchezza.

Vi sono infatti questioni che, per il loro oggetto, sono di interesse esclusivo o quasi esclusivo dei ricercatori italiani e non vi è alcuna buona ragione per ricorrere alla lingua franca; basterebbe pensare a un'infinità di problemi giuridici di nessun interesse per chi non si occupi dell'ordinamento italiano. Così pure vi sono discipline per le quali l'uso della lingua italiana non costituisce una barriera significativa, poiché la comunità internazionale di riferimento conosce l'italiano (qualche esempio: la storia dell'arte medievale, la letteratura latina, le discipline italianistiche, ecc.).

D'altro canto non vi è dubbio che numerosi prodotti della ricerca italiani non circolano come meriterebbero unicamente a causa della lingua in cui sono redatti. Che fare dunque? La questione non può che coinvolgere gli attori del mercato editoriale.

Sono gli editori e coloro che hanno la responsabilità delle riviste che dovrebbero essere indotti, anche attraverso un dialogo e un processo di riflessione condivisa, ad adottare politiche tali da favorire una percentuale crescente di pubblicazioni in lingua inglese, là dove la cosa può avere una ricaduta effettivamente positiva sull'impatto della ricerca.

Senza ricadere in aberrazioni delle quali si è purtroppo parlato come di prossima introduzione, come la redazione di "classifiche di editori" o il computo di un elevato *rejection rate*³³ quale fattore di una positiva valutazione, si dovrebbe fare il possibile per ottenere che l'intero mondo dell'editoria scientifica partecipi a una discussione aperta intorno alla qualità della ricerca, in modo da favorire l'adozione delle pratiche migliori in termini di *peer review* e di accettazione dei contributi.

In questo quadro, occorrerebbe avviare anche una riflessione complessiva e approfondita sul sistema dell'editoria scientifica italiana: infatti, nel nostro Paese vi sono editori che si sostengono attraverso la pubblicazione di prodotti della ricerca finanziati (con fondi della ricerca), senza dunque alcun rischio d'impresa ed editori che affrontano il rischio commerciale della pubblicazione anche ricorrendo a

³³ Il *rejection rate*, un valore assurdo e del tutto non verificabile corrispondente alla percentuale di articoli rifiutati da una rivista, che sarebbe secondo alcuni indicativo della particolare selettività di una sede editoriale, è qui spiritosamente messo alla berlina: <http://www.universalrejection.org/>.

comitati interni di *peer reviewers*, secondo un meccanismo più conforme agli *standards* internazionali.

Questo è un punto critico che va affrontato con molta cautela: infatti è chiaro che l'editore che seleziona i contributi attraverso la propria *peer review* e affronta in tutto o in parte i rischi della pubblicazione, contribuisce a favorire un miglioramento qualitativo della produzione scientifica.

D'altro canto non possiamo, né per le HSS né per le scienze dure affidarci interamente a meccanismi di selezione puramente "di mercato": infatti un editore perseguirà pur sempre il profitto della propria impresa ed è quindi probabile che finiscano per essere rigettati non solo prodotti di cattiva qualità, ma anche prodotti poco spendibili sul mercato. Un problema, ancora una volta, che rischia di essere più rilevante per i settori con un minor numero di ricercatori e con un mercato di conseguenza più ristretto.

Un altro tema strettamente connesso a quelli appena visti è quello dell'*open access*. Allo stato attuale, una parte rilevante del finanziamento pubblico alla ricerca per le HSS si traduce in contributi per la stampa di volumi e pubblicazioni varie. Si potrebbe pertanto perfino contestare la definizione di "finanziamento alla ricerca" visto che in realtà a essere finanziate sono in larga parte le pubblicazioni. Ora, a prescindere dalle considerazioni esposte più sopra, pare assurdo che il danaro pubblico sia speso per pubblicazioni gravate da *copyright* per le quali è sistematicamente escluso l'*open access*. Uno dei maggiori vantaggi del nostro tempo, dal punto di vista della ricerca scientifica, è dato proprio dalla facilità di disseminare pubblicazioni rendendole rapidamente accessibili alla comunità interessata.

Che il finanziamento pubblico contribuisca alla pubblicazione di prodotti della ricerca ad accesso ristretto pare un controsenso. Occorre dunque favorire la pubblicazione *open access* dei risultati della ricerca, come del resto suggerisce anche la Conferenza dei Rettori,³⁴ e adottare tutte le misure atte a scoraggiare pubblicazioni interamente proprietarie, almeno là dove esse vedono la luce grazie al finanziamento pubblico.³⁵

Esistono già numerosi *repositories* di pubblicazioni accademiche, da *academia.edu* ad *arXiv*, fino agli archivi della ricerca dei singoli atenei; sarebbe senz'altro meritorio se ci si adoperasse per rendere liberamente accessibile una quota più vasta possibile dei prodotti della ricerca.³⁶

Per quanto riguarda gli esercizi di valutazione, cosa fare della bibliometria? Come ho cercato di dimostrare vi sono molti elementi che suggeriscono quanto meno cautela nel ricorso a tali strumenti. Nulla vieta di avviare sperimentazioni, purché siano tali e si prestino a essere discusse in modo aperto nelle comunità scientifiche di riferimento. Si potrebbero avviare attività di *testing* che comportino un *mix* di

³⁴ CRUI 2009.

³⁵ Cfr. ALLEA 2012: *Publications should be made openly available online, as soon and as freely as possible, as should also educational resources and software resulting from publicly funded research.*

³⁶ Occorrerà però evitare la proliferazione di editori commerciali *open access*, che si finanziano con il danaro degli autori, un fenomeno già noto all'estero.

strumenti bibliometrici insieme a strumenti differenti, allo stato poco sperimentati, come la *Library Catalog Analysis* (LCA).³⁷

Come si è già avuto modo di dire, nelle HSS gli indici citazionali rischiano di avere una bassa rilevanza statistica, il che significa che essi non restituiscono informazioni significative rispetto all'impatto dei prodotti oggetto di valutazione. Si potrebbe dunque scartare recisamente questi strumenti per sostituirli con la *peer review*. Tuttavia, non si può trascurare il fatto che la *peer review*, al di là delle critiche che si possono muovere allo strumento, è particolarmente costosa. Se pure si può pensare di ricorrere a *peer review* per un esercizio di valutazione nazionale nel quale si chiede ai ricercatori di sottoporre a giudizio i loro prodotti migliori, e dunque un numero limitato – per quanto vasto – di prodotti, la cosa diviene impensabile se si considera, ad esempio, la valutazione interna d'ateneo, che deve considerare l'intera produzione della struttura in un certo arco temporale. Per questo motivo si potrebbe pensare ad un uso combinato di indicatori bibliometrici, LCA, e altri parametri di carattere qualitativo (ad es. recensioni, premi, traduzioni, etc.). In ogni caso dovrebbe trattarsi di sperimentazioni da condursi con cautela e costanti verifiche in modo da giungere nel tempo alla definizione dei migliori strumenti di valutazione.

Nel frattempo, sarebbe comunque bene che gli esercizi di valutazione fossero accompagnati da una articolata campagna informativa intorno alle *best practices*, contribuendo a diffondere fra i ricercatori italiani una diffusa consapevolezza dei problemi della valutazione e della scientometria.

Molto si è discusso circa la revisione dei pari, mettendone anche in dubbio l'affidabilità;³⁸ tuttavia per le HSS sembra allo stato che la valutazione, almeno a livello nazionale, si possa utilmente basare unicamente su di essa, il che peraltro non è in contrasto con quanto accade in alcuni dei maggiori esercizi di valutazione europei. Una *peer review* anche "guidata",³⁹ purché trasparente. Trasparente circa le informazioni trasmesse ai revisori, trasparente circa l'elaborazione delle schede predisposte per la valutazione, trasparente perché si renderanno noti i nomi dei *reviewers ex post*, pur senza specificare quali prodotti ogni singolo *reviewer* ha valutato.

Senza trasparenza, infatti, non può esservi valutazione degna di questo nome.

Vorrei ricordare a questo proposito le giustificate preoccupazioni di Arnold e Fowler riguardo un cattivo uso delle tecniche di valutazione: *the consequences of this unfortunate situation are great. Rewards are wrongly distributed, the scientific literature and enterprise are distorted, and cynicism about them grows. What is to be*

³⁷ Cfr. FAGGIOLANI, SOLIMINE 2012.

³⁸ Cfr. p. es. CECI, PETERS 1982. Da notare che Ceci e Peters suggeriscono che la stessa *peer review* non è distorta tanto da affiliazioni a determinati gruppi accademici o dal *groupthinking*, ma dall'uso dei *rankings*: l'appartenenza di uno studioso a un'istituzione rinomata diviene di per sé il criterio per approvare i risultati della sua attività di ricerca. Si tratta insomma di un'altra manifestazione dell'effetto S. Matteo. Sull'affidabilità della revisione dei pari cfr. BACCINI 2010, p. 52 ss.; GILLIES 2008, p. 13 ss.

³⁹ Come già annuncia di voler fare ANVUR. Cfr. in proposito BANFI 2012.

*done? Just as for scientific research itself, the temptation to embrace simplicity when it seriously compromises accuracy, must be resisted. Scientists who give in to the temptation to suppress data or fiddle with statistics to draw a clearer point are censured. We must bring a similar level of integrity to the evaluation of research products. Administrators, funding agencies, librarians, and others needing such evaluations should just say no to simplistic solutions, and approach important decisions with thoughtfulness, wisdom, and expertise.*⁴⁰

6. Un solo modello di finanziamento?

Alle considerazioni sopra svolte, si lega la questione del modello di finanziamento della ricerca: è infatti dalle scelte operate in questo campo che discende il modo in cui si articola il sistema della valutazione.

In Italia si è cominciato solo di recente ad avviare un vero e proprio esercizio di valutazione del sistema dell'università e della ricerca. Non riflettere sugli aspetti strutturali del sistema solo perché esso è giovane sarebbe un grave errore. Il sistema italiano è strettamente imparentato con quello inglese (RAE, *research assessment exercise*; REF, *research excellence framework*). Un sistema le cui radici rimontano all'era del primo ministro Thatcher. Lo scopo del RAE/REF era quello di assicurare, in una situazione di risorse scarse, la distribuzione più efficiente possibile di queste risorse, promuovendo la loro massima produttività.⁴¹

Va detto che nel contesto italiano, l'esercizio di valutazione è funzionale alla distribuzione di una quota del Fondo di Finanziamento (FFO) definita "premiale", che in realtà nel quadro dei tagli costanti degli ultimi anni premiale non è, poiché non si aggiunge alle risorse necessarie per garantire il buon funzionamento delle strutture, ma ne è in realtà parte.

Siamo sicuri che questo sia il modo giusto di procedere? La VQR è un esercizio dispendioso,⁴² che coinvolge 450 valutatori scelti da ANVUR, i quali dovranno provvedere a scegliere un ampio numero di *reviewers* per procedere alla valutazione di un numero di prodotti che supera ampiamente le 200.000 unità. Le energie sottratte alla ricerca, e le risorse impiegate potrebbero essere superiori ai risultati, ammesso che essi si producano, rendendo l'intero esercizio del tutto antieconomico.⁴³

Vedremo cosa accadrà; tanto più che gli effetti dei forti tagli degli ultimi anni potrebbero manifestarsi già nei prossimi mesi o anni anche nelle statistiche relative alla produzione dei ricercatori italiani.

In ogni caso, vi è un problema più generale: nei paesi che da più tempo si affidano a esercizi di valutazione della qualità della ricerca, sono sempre più numerosi i ricercatori che iniziano a manifestare dubbi intorno ai benefici che tali esercizi producono, rispetto ai costi.

⁴⁰ ARNOLD, FOWLER 2011, p. 437.

⁴¹ GEUNA, MARTIN 2003, p. 280 ss.

⁴² SIRILLI 2012. Le stime di Sirilli sono state contestate da ANVUR, ma va ricordato che il costo del REF, esercizio per certi versi equivalente a quello italiano, è stimato in 100 milioni di sterline (cfr. SIRILLI 2012.2).

⁴³ ABRAMO, CICERO, D'ANGELO 2012, pp. 167- 168.

Non varrebbe la pena di iniziare a riflettere su altre possibili strade, che ci consentano di porci all'avanguardia, anziché ripercorrere acriticamente quanto già tentato altrove?

In un suo recente scritto B. R. Martin osserva come un esercizio di valutazione, perché valga la pena intraprenderlo, dovrebbe avere costi contenuti (inferiori all'1% del costo delle attività valutate) e assicurare benefici nettamente superiori ai costi.⁴⁴ Il REF (già RAE) britannico sembra mostrare costi crescenti e benefici decrescenti: del resto ci si trova dinnanzi a una sorta di trappola. Un sistema di valutazione poco costoso e molto snello rischia di non catturare appieno la "fotografia" della situazione. Un sistema che sia pensato per svolgere al meglio questa attività rischia però di divenire assurdamente costoso.⁴⁵

Tornando al caso italiano, sarebbe bene non considerare unicamente i costi complessivi della VQR (che peraltro è pressoché impossibile stimare accuratamente), ma considerarli anche in relazione alle aree disciplinari.

Infatti, alcune di esse (proprio le 10, 11, 12, 13, 14) non necessitano di fondi in quantità paragonabili a quanto deve essere assegnato alle "scienze dure". I fondi sono infatti prevalentemente assorbiti dai sistemi bibliotecari e da spese per pubblicazioni, non certo per attrezzatura di laboratorio e spese per personale. Il che rende, almeno per queste aree, ancora meno sensato il rapporto fra i costi e i benefici di un esercizio di valutazione così congegnato.

Si sa che esiste un altro modello di finanziamento: lo Stato assicura le attività di funzionamento ordinario e la didattica, limitando il finanziamento alla ricerca unicamente a progetti (*grants*), questi soli oggetto di valutazione. E' il modello statunitense. Un modello che se presenta indubbi vantaggi in termini di costi e snellezza del sistema, pone comunque alcuni problemi rilevanti.

Ne segnalo tre: la trasformazione degli scienziati in compilatori di bandi a tempo pieno, a danno della qualità e della quantità della ricerca medesima. In secondo luogo, i finanziamenti a progetto, specie se concentrati su progetti di vasta entità, tendono a sopprimere la ricerca *curiosity driven*, che non solo è dominante nell'ambito delle HSS, ma ha una funzione essenziale anche per il progresso delle scienze dure, tanto che si rende necessario introdurre dei correttivi, attraverso finanziamenti "a pioggia" o un certo numero, ragionevolmente ampio, di piccoli *grants*. Infine vi è il rischio che il sistema favorisca l'effetto S. Matteo allocando la stragrande maggioranza dei fondi a favore di poche università ritenute a torto o a ragione "di élite", il che può avere effetti non positivi (il pluralismo e la ricchezza della ricerca non beneficiano certo della concentrazione).⁴⁶

Il dibattito internazionale è vivace non solo sul "come" valutare, ma anche sul "chi e cosa" valutare. Si discute sull'opportunità di adottare diversi modelli di

⁴⁴ MARTIN 2011, p. 248; cfr. anche GEUNA, MARTIN 2003, p. 300 ss.

⁴⁵ MARTIN 2011, p. 251 ss.

⁴⁶ Su questi punti cfr. IOANNIDIS 2001 e TATSONI, VAVVA, IOANNIDIS 2010; GEUNA, MARTIN 2003, p. 296.

finanziamento (a pioggia, a sorteggio, a progetto, centralizzato), o loro combinazioni.

E' bene che la comunità scientifica italiana, e tutti gli attori del sistema non rimangano estranei alla discussione, e che si avvii una riflessione assidua su questi temi, al fine di assicurare gli esiti migliori per il sistema italiano dell'università e della ricerca.

7. Appendice: aspetti critici dell'esercizio italiano.

A conclusione di queste pagine, alcune brevi osservazioni riguardo le anomalie che allo stato si sono potute riscontrare nel corrente esercizio di valutazione condotto da ANVUR.

a. Rankings di riviste.

Si è già scritto molto su ANVUR e sui difetti "di progettazione" dell'Agenzia, che non nasce terza, ma come *executive agency*.⁴⁷ I primi passi del VQR 2004-2010 non sono incoraggianti. Il dialogo fra valutatori e valutati è stato molto difficile, anche a causa di alcuni aspetti discutibili della metodologia di ANVUR. Gli strumenti bibliometrici non sono sempre stati approntati in modo impeccabile. Sono state fatte scelte molto discutibili per un esercizio strutturato come il VQR, quale il ricorso a liste di riviste per coadiuvare il lavoro dei valutatori, nonostante esercizi di valutazione stranieri ormai consolidati preferiscano evitare di utilizzare strumenti di questa natura.⁴⁸ Al contempo, si è cercato di tranquillizzare i ricercatori spiegando loro che le liste, per buona parte delle HSS, saranno fornite come "materiale aggiuntivo" ai revisori dei singoli prodotti della ricerca, senza che vi sia alcun automatismo che produca il trasferimento del valore del contenitore sul contenuto.

Eppure è facile prevedere che i revisori, se gravati di una certa mole di prodotti, potranno essere indotti ad agire in tal modo.

Gli stessi GEV in più punti dei loro documenti programmatici tradiscono forti ambiguità riguardo al valore delle classifiche.⁴⁹

Inoltre, le liste di riviste prodotte da ANVUR presentano problemi gravi sia per le aree bibliometriche che per le HSS. Quanto alla creazione dei *rankings* per le HSS è stato reso noto unicamente il processo, non i criteri, che pure si sarebbero potuti stabilire in modo oggettivo.

⁴⁷ Cfr. PINELLI 2011.

⁴⁸ Per una feroce e documentata critica dei rankings, cfr. WILLMOTT 2011.

⁴⁹ Scrive ad es. il GEV 11: "i saggi pubblicati su riviste internazionali di fascia A" saranno inviati a un solo revisore "per tener conto del fatto che essi sono già passati attraverso filtri molto rigorosi e premiare [...] lo sforzo di internazionalizzazione"; "in caso di divergenza fra il giudizio dei revisori [...] e la classificazione delle riviste [...] il primo prevarrà se concorde; se uno solo dei revisori produce una valutazione che si discosta dalla classificazione [...] la divergenza viene trattata come una fra revisori". In determinate circostanze, dunque, la classe della rivista è equiparabile al giudizio di un peer. Cfr. BANFI 2012; BANFI 2012.2. Il testo del documento del GEV 11 è reperibile qui: http://www.anvur.org/sites/anvur-miur/files/gev_documenti/gev11_allegati.zip.

Poiché questi criteri non sono noti, si potrebbe perfino presumere che essi non esistano neppure. Il che non è bene poiché non consente di verificare la correttezza delle procedure seguite. In effetti sappiamo che si è provveduto a chiedere elenchi alle società disciplinari, corredati di una valutazione relativa ad ogni singola rivista (A, B, C, D). Questi elenchi sono stati rivisti da esperti anonimi (perché non rivelarne *ex post* l'identità?) e infine validati (non si sa se con ulteriori modifiche) dai singoli GEV. Una procedura non del tutto trasparente, che ha dato luogo a fenomeni per più versi criticabili.

Ne cito solo tre, a titolo di esempio:

1. Vi sono riviste che cambiano *rating* a seconda che siano valutate con riferimento all'una o all'altra disciplina all'interno di una stessa area.

Questo fenomeno è comune nelle aree delle scienze dure. In effetti, se si procede a classificare le riviste su base bibliometrica per percentili, è evidente che al variare del settore disciplinare varierà la distribuzione delle citazioni e dunque potrà eventualmente mutare la valutazione della rivista. Questo non avviene per le HSS, dove le riviste sono classificate non su base bibliometrica. In buona sostanza, in assenza di informazioni bibliometriche, non si può che classificare le riviste sulla base di *indicatori oggettivi di qualità*, sul genere di quelli che si possono ritrovare nella lista iberoamericana di LATINDEX.⁵⁰ Pertanto, se si fossero adottati criteri oggettivi per classificare le riviste delle HSS (p.es. internazionalizzazione, composizione del comitato scientifico, procedure di *peer review* verificabili e *double blind*, presenza nelle biblioteche nazionali e internazionali, e così via) il fenomeno del cambio di classe al mutare del settore di appartenenza del contributo non avrebbe dovuto prodursi: gli indicatori di qualità rimangono infatti gli stessi. Peraltro, va rilevato che il bando VQR prevede l'attribuzione dei prodotti della ricerca secondo classi di merito così articolate: A eccellente (20% superiore della scala di valore), B buono (segmento 60-80% della scala di valore), C accettabile (segmento 50-60%), D limitato (segmento del 50% inferiore).

Questa distribuzione non è stata rispettata nella stesura delle liste delle riviste di HSS, e di conseguenza i cambi di classe di una stessa rivista al variare del settore del contributo pubblicato, benché all'interno di una medesima area disciplinare, non sono giustificabili neppure con l'esigenza di rispettare la distribuzione percentile stabilita dal bando VQR.

Ciò conferma che le attribuzioni di *ratings* nelle HSS sono avvenute su di una base per così dire "discrezionale". Del resto lo stesso vice coordinatore dell'esercizio di valutazione deve ammettere che in difetto di altri strumenti (ma davvero non ce n'erano?) si è fatto ricorso a un criterio (la "reputazione") assolutamente soggettivo. E' davvero solo la metrica citazionale a essere "oggettiva"? Scrive Bonaccorsi: "nell'esercizio dei GEV non è stata adottata una definizione formale delle categorie, ma si è fatto riferimento ad una definizione sintetica di reputazione. Essa include una molteplicità di dimensioni. In sede di aggiornamento dei rating occorrerà standardizzare la definizione delle categorie, in modo da fornire ad ogni valutatore

⁵⁰ http://www.latindex.unam.mx/documentos/revistas_imp.html.

un insieme di informazioni omogeneo e ben chiaro dal punto di vista semantico.”⁵¹ Pare dunque di capire che le graduatorie siano state costruite a prescindere da “un insieme di informazioni omogeneo e ben chiaro dal punto di vista semantico”. Come afferma il gruppo di esperti della valutazione dell’area 11 (scienze storiche, filosofiche, pedagogiche e psicologiche), nella futura revisione dei *rankings* “si potranno allora utilizzare criteri oggettivi che non era possibile applicare retrospettivamente in modo meccanico (sta qui la ragione dell’ampio ricorso fatto in questi elenchi alla reputazione così come giudicata dalla comunità scientifica)”.

In effetti, vi sono riviste poste nelle fasce più elevate che sono prive di *peer review*, benché si tratti di un criterio discriminante per la qualità di una pubblicazione a livello internazionale, e che gli stessi valutatori annunciano di voler considerare come determinante.⁵² Così pure sono classificati come riviste prodotti editoriali che non hanno i requisiti per essere considerati tali.⁵³

Le liste sono state redatte, dunque, sulla base di concertazioni che hanno coinvolto società disciplinari e non meglio identificati soggetti anonimi senza tenere in conto parametri oggettivi, trasparenti e verificabili. Attribuire “bollini di qualità” per gli anni 2004-2010 applicando retrospettivamente criteri non ancora pienamente condivisi dalle comunità scientifiche sarebbe stata senz’altro una stortura. Optare per una concertazione opaca non pare tuttavia la migliore delle soluzioni possibili.

2. Si usa difendere i *rankings* di riviste affermando che essi hanno un effetto pedagogico e giovano a orientare gli studiosi (specie i giovani) verso le migliori sedi di pubblicazione, e che in tal modo essi favoriscono indirettamente la competizione. Viene allora da chiedersi per quale motivo riviste di primaria importanza internazionale sono esclusi dalle graduatorie (una per tutte, pensando alle scienze giuridiche: la *Harvard Law Review*)⁵⁴.

Che fine fa, in tal caso, l’effetto “pedagogico” dei *rankings*?⁵⁵

3. l’Agenzia ha informato la comunità accademica della possibilità di provvedere a una revisione del *rating* assegnato a una singola rivista. Senza entrare nei dettagli della procedura, merita di essere segnalato come unicamente il direttore della rivista o il suo editore siano legittimati a chiedere la revisione del *rating*. Se questa scelta può apparire per certi versi comprensibile, va però osservato che essa avrà la conseguenza di rimettere in discussione unicamente il *rating* di riviste designate come non eccellenti: l’attribuzione della classe A si configura pertanto come un giudizio inappellabile. Infatti nessun direttore o editore ne chiederà mai la revisione, né questa possibilità è concessa a eventuali controinteressati. Il che, ancora una volta, suscita dubbi sulla *fairness* della procedura: si ammette la possibilità di sottovalutazioni, ma non quella di sopravvalutazioni. Il che non pare ragionevole. Se si fosse ottenuto un largo consenso sui *rankings*, sulla base di criteri chiari e

⁵¹ BONACCORSI 2012, p. 13.

⁵² BONACCORSI 2012, p. 14 ss.

⁵³ Mancanza di ISSN, periodicità irregolare, etc.

⁵⁴ Ma non solo, cfr. BANFI 2012.

⁵⁵ A proposito della pedagogia dei *rankings*, cfr. BACCINI, BANFI 2012.

verificabili e di una consultazione ampia, approfondita e trasparente delle comunità scientifiche, prevedendo al contempo una revisione delle liste con cadenza annuale, si sarebbe potuto evitare di escogitare meccanismi di revisione che finiscono per confermare l'inattendibilità dell'intero sistema di *rankings*. In effetti, l'impressione è che la procedura di revisione sia un tentativo di prevenire possibili contenziosi giudiziari, piuttosto che uno strumento per assicurare la correttezza dell'intera procedura.

b. *Impiego di database inadeguati e altre anomalie.*

L'area 13 (scienze economiche e statistiche) sceglie di utilizzare Google Scholar come strumento per la costruzione di parte delle proprie liste, almeno relativamente all'h-index;⁵⁶ altre aree⁵⁷ scientifico-disciplinari useranno Google Scholar come strumento di *testing* bibliometrico. Eppure c'è un'ampia letteratura che dimostra come si tratti di una base dati ancora del tutto inaffidabile, fra l'altro, come già si è detto, aperta a manipolazioni.

Qualcuno potrà fare spallucce, sottolineando come qui si tratti di valutazione di strutture e non di singoli e argomentando che i grandi numeri (ma *quanto* grandi?) consentono di minimizzare e rendere irrilevante il margine di errore.

Non è così; come io stesso ho avuto modo di dimostrare in forma scherzosa, un uso disinvolto di database non adatti allo scopo è in grado di azzerare la misurazione non solo dell'h-index, ma della stessa produttività per interi settori disciplinari.⁵⁸

A questo proposito, va ribadito con forza che da cattive basi dati non possono che derivare cattive valutazioni, peraltro esposte al rischio di contenzioso.⁵⁹

Nessun serio esercizio di valutazione può prescindere dall'esistenza di basi dati attendibili, affidabili e trasparenti.

Inoltre, pretendere di rivedere in futuro le classifiche di riviste adottando come criterio, fra gli altri, l'esito della valutazione in corso⁶⁰ mi pare un paralogismo davvero straordinario. Rivedere un *ranking* sulla base degli esiti prodotti almeno in parte dallo stesso *ranking*? Sarà anche un criterio oggettivo, ma pare davvero poco opportuno adottarlo.

Ancora, non è chiaro perché all'interno delle HSS si creino disomogeneità apparentemente inspiegabili anche per quanto riguarda i prodotti oggetto di valutazione. L'area 10 (scienze dell'antichità, filologico-letterarie e storico-artistiche) ritiene che anche le voci in dizionario e enciclopedia possano essere oggetto di valutazione, al pari di prefazioni e postfazioni, edizioni di fonti inedite e concordanze.⁶¹ Non entro nel merito della scelta che può essere anche ragionevole, ma rilevo che questi prodotti sono frequenti anche per le aree 11 e 12. Per quale motivo risultano valutabili solo nell'area 10? Scelte di questo genere non giovano alla credibilità dell'intero esercizio di valutazione.

⁵⁶ BACCINI 2012.1.

⁵⁷ Cfr. ad es. i documenti del GEV 12 sul sito di Anvur (www.anvur.org).

⁵⁸ BANFI 2012.1.

⁵⁹ Cfr. GALIMBERTI 2012.

⁶⁰ BONACCORSI 2012, p. 16.

⁶¹ Comunicazione ANVUR del 24 aprile 2012, peraltro di molto posteriore alla pubblicazione dei criteri di valutazione.

c. *Valutare o condizionare?*

ANVUR si propone per il futuro di contribuire allo sviluppo delle scienze umane e sociali attraverso l'elaborazione di nuovi strumenti di analisi dedicati a uno dei principali prodotti della ricerca per queste aree: la monografia. E' noto infatti che le monografie sfuggono per ora all'analisi bibliometrica, e poiché esse costituiscono una larga parte della produzione scientifica di numerosi settori scientifico-disciplinari delle HSS, ciò rende pressoché impossibile condurre un'analisi bibliometrica seria di questi settori. D'altro canto la stessa *peer review* di monografie è particolarmente onerosa.⁶²

Sarebbe già possibile computare le citazioni di una monografia, purché esse siano contenute in riviste indicizzate nei database WoS e Scopus. Va rilevato però che solo poche, se non pochissime riviste delle HSS, specialmente per le aree 10, 11 e 12 sono indicizzate, il che non consentirebbe comunque di ottenere dati attendibili, a prescindere da tutte le considerazioni svolte in precedenza.

D'altra parte, approntare sistemi di *data mining* per le monografie delle HSS (risultano esservi numerose attività sperimentali a tal fine, avviate anche, ma non solo, presso i grandi oligopolisti delle banche dati scientifiche), ancora una volta condurrebbe a risultati solo parzialmente attendibili, poiché si estrarrebbero (faticosamente) le citazioni da monografie verso monografie e da monografie verso articoli di rivista. Per ottenere risultati utili occorrerebbe interfacciare questi risultati con quelli provenienti dall'analisi citazionale – incompleta, come si è detto - delle riviste e solo di quelle. Un compito tutt'altro che semplice. Inoltre, tutto ciò ha senso solo se si decide di applicare in modo più o meno meccanico il paradigma dell'analisi citazionale alle HSS; la futura "analisi citazionale delle monografie" non risolve però in alcun modo i problemi sopra sollevati circa il metodo e lo stile della ricerca di alcune delle HSS.

A che pro, dunque, perseguire questa via, che può rivelarsi anche onerosa?

Va anche rilevato che soluzioni alternative, come quella da più parti caldeggiata della costruzione di graduatorie di editori,⁶³ potrebbero avere risultati distorsivi non secondari sul mercato editoriale, tali da consigliare la massima cautela, anche in considerazione del contenzioso che ne potrebbe derivare. In realtà, sarebbe forse bene evitare certe forme di feticismo bibliometrico,⁶⁴ e modulare con cura gli strumenti di analisi rispetto ai fini che si vogliono ottenere e a ciò che si intende valutare.

Ma c'è un problema ancora più complesso che deve essere affrontato. L'Agenzia italiana di valutazione si muove all'interno di uno spazio in cui non incontra contrappesi: né, per ragioni di carattere contingente, da parte del Ministero, né da parte del Consiglio Universitario Nazionale, uscito alquanto depotenziato dal

⁶² GIMÉNEZ-TOLEDO, ROMÁN-ROMÁN 2009, p. 202.

⁶³ GIMÉNEZ-TOLEDO, ROMÁN-ROMÁN 2009, pp. 201-213

⁶⁴ WILLMOTT 2011, p. 16: *For journals list fetishists, the pleasure of scholarship does not reside so much in undertaking the research as in moulding and squeezing it into a form – in terms of style and content - that we anticipate will render it acceptable to publication in the targeted top ranked journal.*

processo di riforma, né da parte delle comunità scientifiche spesso strutturalmente deboli.

In questo quadro l'agenzia, alla quale la l.240/2011 assegna una serie di compiti inquadrabili solo in un concetto molto lato di valutazione (un ruolo le è assegnato perfino nel caso del dissesto finanziario degli atenei)⁶⁵ ha assunto un atteggiamento fortemente dirigista, come traspare anche dalle dichiarazioni rese tempo fa dal responsabile della valutazione a un quotidiano, secondo le quali si andrà verso un ridimensionamento dei corsi di dottorato, chiusura di corsi di laurea e perfino di interi atenei.⁶⁶

Penso che un'agenzia ben congegnata dovrebbe impegnarsi per fertilizzare il terreno in modo che questo sia il più produttivo possibile, senza intervenire direttamente scegliendo quale pianta far crescere e quale estirpare: ciò andrebbe oltre la valutazione, per investire l'ambito delle *policies*.

Le scelte di ANVUR paiono però indirizzarsi proprio in questo senso. Ad esempio, l'Agenzia non intende solo diffondere informazioni e favorire da parte delle riviste delle HSS l'ingresso nelle grandi basi dati internazionali, bensì scegliere essa stessa *quali* riviste "eccellenti" (pare di capire una rosa ristretta fra quelle classificate come di fascia A) "aiutare" direttamente ad entrare in tale circuito;⁶⁷ si tratta di un interventismo che tradisce una concezione meccanicistica della ricerca, per cui essa può essere pianificata e orientata da parte di un organo centralizzato.

Resto dell'idea che la ricerca non possa essere né pianificata né orientata; che si tratti di un sistema troppo delicato per essere manovrato unicamente con il metodo aziendalistico di incentivi e disincentivi.

La ricerca è fatta soprattutto di libertà e pluralismo, senza i quali non cresce, e chi la finanzia (o la valuta ai fini del finanziamento) deve stare ben attento a non sopprimere tali caratteristiche sulla base di una malintesa idea di produttività, che rischia – nel migliore dei casi – di rivelarsi efficace solo nel breve e in termini meramente numerici.

E' evidente che l'esercizio di valutazione ormai in corso è avviato ed è molto difficile intervenire con correttivi che pure sarebbero opportuni. Per il futuro si può e si deve, invece, pensare a delle modifiche migliorative.

⁶⁵ DLGS 199 27/10/2011 Disciplina del dissesto finanziario delle università, art. 19.

⁶⁶ S. Benedetto su La Repubblica del 4.2.2012, p. 50.

⁶⁷ BONACCORSI 2012,, p. 7: "In questa linea di attività l'ANVUR intende sostenere la candidatura di un consistente gruppo di riviste italiane alla indicizzazione presso ISI e Scopus. Appare ragionevole che lo sforzo maggiore sia indirizzato alle riviste di fascia A, definite secondo il rating proposto dai GEV della VQR. Non è esclusa la estensione alle riviste di fascia B, con opportune verifiche."

Bibliografia

. ABRAMO, CICERO, D'ANGELO 2012: G. Abramo, T. Cicero, C. A. D'Angelo, *The dispersion of research performance within and between universities as a potential indicator of the competitive intensity in higher education systems*, Journal of Informetrics 6 (2012), pp. 155-168.

.ADAMS 2007: J. Adams, *Bibliometrics, assessment and UK research*, Serials 20(3), November 2007 (<http://uksg.metapress.com/content/10yj8m7h3e4ljcmw/fulltext.pdf>)

. ADLER, EWING, TAYLOR 2008: R. Adler, J. Ewing, P. Taylor, Joint Committee on Quantitative Assessment of Research, *Citation Statistics, A 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)* (<http://www.roars.it/online/wp-content/uploads/2012/01/CitationStatistics.pdf>)

. ALLEA 2012: Open Science for the 21st century. *A declaration of ALL European Academies presented at a special session with Mme Neelie Kroes, Vice-President of the European Commission, and Commissioner in charge of the Digital Agenda on occasion of the ALLEA General Assembly held at Accademia Nazionale dei Lincei, Rome, on 11-12 April 2012* (www.allea.org)

. ARNOLD, FOWLER 2011: D. N. Arnold, K. K. Fowler, *Nefarious Numbers*, Notices of the AMS 58: 434-437, March 2011 -ArXiv, 1 Oct 2010.

.BACCINI 2010: A. Baccini, *Valutare la ricerca scientifica. Uso e abuso degli indicatori bibliometrici*, Bologna 2010.

.BACCINI 2011: A. Baccini, *La valutazione della ricerca e i suoi strumenti*, Annali di Storia delle Università Italiane, 15/2011, pp. 9-28.

.BACCINI 2012.1: A. Baccini, *Tu chiamala, se vuoi, valutazione*, Roars Review I/2012 (<http://www.roars.it/online/?p=7498>).

.BACCINI 2012.2: A. Baccini, *Il VQR e il vino a due stadi* (<http://www.roars.it/online/?p=5164>)

.BACCINI, BANFI 2012: A. Baccini e A. Banfi, *Limiti e pericoli delle classifiche di riviste*, Roars Review II/2012 (<http://www.roars.it/online/?p=7914>)

.BANFI 2012: A. Banfi, *Osservazioni sulla valutazione della ricerca nelle scienze umane e sociali*, ASTRID Rassegna 155 (6/2012). (<http://www.irpa.eu/wp-content/uploads/2012/03/Valutazione-HSS.pdf>)

. BANFI 2012.1: A. Banfi, *Scholar Search: all'inferno andata e ritorno*, Roars Review I/2012 (<http://www.roars.it/online/?p=7679>)

. BANFI 2012.2: A. Banfi, *Aspetti critici dell'uso di rankings di riviste nelle scienze umane*, ASTRID Rassegna 153 (4/2012). (<http://www.irpa.eu/wp-content/uploads/2012/02/rankings.pdf>)

.BONACCORSI 2012: A. Bonaccorsi, *Potenzialità e limiti dell'analisi bibliometrica nelle aree umanistiche e sociali, verso un programma di lavoro*. (http://www.anvur.org/sites/anvur-miur/files/valutazione_aree_umanistiche_e_sociali.pdf)

- . CASSELLA, BOZZARELLI 2011: M. Cassella, O. Bozzarelli, *Nuovi scenari per la valutazione della ricerca tra indicatori bibliometrici citazionali e metriche alternative nel contesto digitale*, Biblioteche oggi – marzo 2011, pp. 66-78.
(<http://www.roars.it/online/wp-content/uploads/2012/01/Nuovi-scenari-per-la-valutazione.pdf>)
- .CECI, PETERS 1982: S. J. Ceci, D. P. Peters, *Peer review: a study of Reliability*, *Change* 14, 6 (sep. 1982), pp. 44-48.
- . CRUI 2009: Crui, Commissione Biblioteche, Gruppo Open Access, *L'open access e la valutazione dei prodotti della ricerca scientifica. Raccomandazioni*.
(<http://www.cruir.it/HomePage.aspx?ref=1782#>)
- .DE NICOLAIO 2011: G. De Nicolaio, *I numeri tossici che minacciano la scienza, trucchi scandali e pericoli degli indicatori bibliometrici*.
(<http://www.roars.it/online/?p=339>)
- .DE NICOLAIO 2012: G. De Nicolaio, *Università: ciò che Bisin e De Nicola non sanno (o fingono di non sapere)*
(<http://www.roars.it/online/?p=7092>)
- .ERNST 2010: R.R. Ernst, *The Follies of Citation Indices and Academic Ranking Lists. A Brief Commentary to Bibliometrics as Weapons of Mass Citation*, *CHIMIA* 2010, 64, No. 1/2 p. 90.
- .FAGGIOLANI, SOLIMINE 2012: C. Faggiolani, G. Solimine, *LCA (Library Catalog Analysis) e la valutazione della ricerca umanistica*
(<http://www.roars.it/online/?p=6735>)
- . GALIMBERTI 2012: P. Galimberti, *Qualità e quantità: stato dell'arte della valutazione della ricerca nelle scienze umane in Italia*, *Jlis.it* 3 n.1 (giugno 2012), 5617-25.
- . GEUNA, MARTIN 2003: A. Geuna, B. R. Martin, *University Research Evaluation and Funding: an International Comparison*, *Minerva* 41 (2003), 277–304.
- . GILLIES 2008: D. Gillies, *How should research be organized?*, London 2008.
- . GIMÉNEZ-TOLEDO, ROMÁN-ROMÁN 2009: E. Giménez-Toledo, A. Román-Román, *Assessment of humanities and social sciences monographs through their publishers: a review and a study towards a model of evaluation*, *Research Evaluation*, 18(3), September 2009, pp. 201-213.
- . GRANDJEAN, ERIKSEN, ELLEGAARD, WALLIN 2011: P. Grandjean, M. L. Eriksen, O. Ellegaard, J. A. Wallin, *The Matthew effect in environmental science publication: A bibliometric analysis of chemical substances in journal articles*, *Environmental Health* 2011.
(<http://www.ehjournal.net/content/10/1/96>)
- .IOANNIDIS 2011: John P. A. Ioannidis, *"More Time for Research: Fund People Not Projects."* *Nature* 477, no. 7366 (2011), pp. 529-31
- . JENSEN, ROUQUIER, CROISSANT 2008: P. Jensen, E.B. Rouquier, V. Croissant, *Testing bibliometric indicators by their prediction of scientists promotion*, *Scientometrics* 78.3 (2009), arXiv:0811.0237v1.
- . LABBÉ 2010: C. Labbé, *Ike Antkare, one of the great stars in the scientific firmament*, *ISSI Newsletter* 6,2 pp. 48-52.

- . MARTIN 2011: B.R. Martin, *The Research Excellence Framework and the 'impact agenda': are we creating a Frankenstein monster?*, Research Evaluation sept. 2011, pp. 247-254.
- .MOED 2010: H.F. Moed, *Citation Analysis in Research Evaluation*, Dordrecht 2010.
- . MOLINIÉ, BODENHAUSEN 2010: A. Molinié, G. Bodenhausen, *Bibliometrics as Weapons of Mass Citation*, CHIMIA 2010, 64, No. 1/2, pp. 78-89.
- . PRATELLI, BACCINI, BARABESI, MARCHESELLI 2011: L. Pratelli, A. Baccini, L. Barabesi, M. Marcheselli, *Statistical analysis of the Hirsch Index*, Scandinavian Journal of Statistics (forthcoming), arXiv:1102.2701v1.
- .QUALITY INDICATORS 2011: Royal Netherlands Academy of Arts and Sciences. Interim report by the Committee on Quality Indicators in the Humanities. *Quality Indicators for Research in the Humanities*, 2011.
(www.knaw.nl)
- .RADICCHI, FORTUNATO, CASTELLANO 2008: F. Radicchi, S. Fortunato, C. Castellano, *Universality of citation distributions: Toward an objective measure of scientific impact*, PNAS November 11, 2008 vol. 105 no. 45, pp. 17268–17272.
- . STORBECK 2012: O. Storbeck, *Coerced citations and manipulated impact factors – the dirty tricks of academic journals*.
(<http://economicsintelligence.com/2012/02/25/coerced-citations-and-manipulated-impact-factors-the-dirty-tricks-of-academic-journals/>)
- . THOMSON CORP. 2002: *Journal self-citation in the Journal Citation Reports® – Science Edition* (2002): A Citation Study from The Thomson Corporation
(<http://www.elyadal.org/seminerler/ssci/docs/selfcitations.pdf>)
- . MARTIN 2011: B.R. Martin, *The Research Excellence Framework and the 'impact agenda': are we creating a Frankenstein monster?*, Research Evaluation sept. 2011, pp. 247-254.
- .PINELLI 2011: C. Pinelli, *Autonomia universitaria, libertà della scienza e valutazione dell'attività scientifica*, Rivista AIC 3/2011.
(<http://www.roars.it/online/wp-content/uploads/2012/02/Pinelli-Autonomia-libertà-valutazione.pdf>)
- .SIRILLI 2012: G. Sirilli, *Si può stimare che la VQR costerà 300 milioni di euro – e a pagarli sarà l'università*
(<http://www.roars.it/online/?p=6885>)
- .SIRILLI 2012.2: G. Sirilli, *Se il vino non è sempre buono*, Roars Review VII.2012
(<http://www.roars.it/online/?p=9083>)
- . TATSIONI, VAVVA, IOANNIDIS 2010: A. Tatsioni, E. Vavva, J. P. A. Ioannidis. *Sources of Funding for Nobel Prize-Winning Work: Public or Private?* The FASEB Journal 24, no.5 (May 1, 2010), pp. 1335-39.
- .UNIVERSITIES UK, *Bibliometrics: Universities UK, The use of bibliometrics to measure research quality in UK higher education institutions*,
(<http://www.universitiesuk.ac.uk/Publications/Documents/bibliometrics.pdf>)
- .VAN RAAN 2006: A. F. J. van Raan, *Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups*, Scientometrics, Vol. 67, No. 3 (2006) 491–502.

<http://sci2s.ugr.es/hindex/pdf/VanRaan2006.pdf>)

. WILHITE, FONG 2012: A. W. Wilhite, E. A. Fong, *Supporting Online Material for Coercive Citation in Academic Publishing*, Science 335, 542 (2012)
(<http://www.sciencemag.org/content/suppl/2012/02/01/335.6068.542.DC1/1212540.Wilhite.SOM.pdf>)

.WILLMOTT 2011: H. Willmott, *Journal List Fetishism and the Perversion of Scholarship: Reactivity and the ABS List* (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1753627)

.ZITT 2005: M. Zitt, *Facing Diversity of Science: A Challenge for Bibliometric Indicators*, MEASUREMENT, 3(1) 2005, pp. 38-49.