

Valutazione e scienze umane: limiti delle attuali metodologie e prospettive future

Paola Galimberti (Università degli Studi di Milano)

“an effective system is not simple and a limited system is not effective”

(KNAW 2011)

Abstract

Con la piena operatività della Agenzia nazionale della Valutazione (ANVUR) hanno preso avvio e si sono intensificate in Italia una serie di attività (e di discussioni) relative alla misurazione e alla valutazione della ricerca.

Gli scopi di queste attività ai vari livelli (sistema nazionale, atenei, dipartimenti, gruppi, singoli) sono molteplici e una certa confusione regna rispetto a che utilizzo destinare le informazioni disponibili. Infatti, da un lato, si conferma l'esigenza di verificare i risultati ottenuti a fronte dei finanziamenti stanziati per poter meglio decidere come allocare le risorse (finanziarie ed umane) per una gestione più efficace delle politiche della ricerca, mentre, dall'altro, la necessità di meccanismi di reclutamento trasparenti, basati su criteri oggettivi e verificabili, costituisce da sempre un tema che appassiona gli accademici e genera reazioni e considerazioni nell'opinione pubblica, anche a fronte di scandali e irregolarità.

Uno dei problemi classici è come fare analisi quantitative (e o qualitative) che permettano il confronto fra aree disciplinari molto diverse fra loro. Mentre per le scienze dure sono stati sviluppati e affinati a livello internazionale modelli e indicatori condivisi sufficientemente robusti e tendenzialmente (quasi sempre) accettati dalle comunità scientifiche, per le scienze umane spesso non si riesce ad andare al di là della mera conta dei lavori di ricerca.

Il nostro paese sconta oltre a un ritardo significativo nello sviluppo di modelli di valutazione della ricerca¹, per le aree umanistiche anche il fatto di parlare una lingua diversa dall'inglese per cui la ricerca di importanza nazionale tende ad esprimersi in italiano (si pensi alla giurisprudenza, alla filologia, alla letteratura delle varie epoche, ad alcune aree delle scienze sociali, all'economia aziendale)

In questo contesto, segnato da forti limiti rispetto alla disponibilità di informazioni e di modelli di riferimento si collocano una serie di studi italiani pubblicati di recente la cui ratio comune è quella della applicabilità di modelli sviluppati per le scienze dure alle scienze umane una volta che il problema degli strumenti a disposizione sia stato risolto². Sembra che l'utilizzo della bibliometria o comunque di indicatori quantitativi possa così diventare la soluzione alla attuale assenza di indicatori per le scienze umane e un buon sostituto (o meglio un complemento) della peer review. In questo testo si cercherà di dimostrare come la complessità delle aree umanistiche sia (ancor più che nelle scienze dure) difficilmente riducibile, e come alla comprensione della stessa concorrano molti elementi alcuni dei quali si traducono con difficoltà in linguaggio sintetico, come la taglia unica poco si adatta a queste discipline e che solo un modello che tenga conto della specificità della ricerca umanistica, dei diversi canali di comunicazione e dei loro target, del loro uso e delle modalità di riconoscimento del valore scientifico possa fornire un quadro attendibile e utilizzabile.

¹ <http://www.slideshare.net/giuseppedn/baccini-latecomer-new-roars>

² Si veda in particolare il testo di A. Bonaccorsi *Potenzialità e limiti dell'analisi bibliometrica nelle aree umanistiche e sociali. Verso un programma di lavoro* (marzo 2012)

Premessa

Garantire criteri confrontabili nella distribuzione dei fondi per la ricerca, senza ricorrere a una ripartizione preventiva per aree tematiche, costituisce un tema che riguarda chi programma e chi gestisce la ricerca a tutti i livelli. Associare questi criteri all'analisi dei risultati – precedentemente - ottenuti è sicuramente lo strumento più utilizzato, sia se applicato su parametri documentabili come nel caso della bibliometria, sia se affidato a giudizi qualitativi come nella peer review. Le cosiddette scienze dure oltre ad assorbire nei paesi OECD il 70/80% dei finanziamenti per la ricerca³, fissano lo scenario di riferimento, anche per le modalità pluriennali di impiego delle risorse e per la presenza in molti casi di rilevanti investimenti strumentali che presuppongono una continuità nel tempo e una valorizzazione di quanto avviato. Sono dunque queste aree a determinare i metodi, gli indicatori, le banche dati utilizzate per la valutazione e il loro contenuto. Sono le scienze dure, che si esprimono tendenzialmente in inglese e, in molti casi, in articoli con molti coautori, per comprendere i diversi ruoli e contributi, pubblicati su riviste con IF, a imporre il modello per la costruzione di indicatori qualitativi e quantitativi, basati sull'IF e sul numero di citazioni ricevute con l'impiego di criteri di normalizzazione più o meno spinti. Con questi indicatori si cerca di restituire in maniera fedele una proxy della qualità e /o dell'impatto di una ricerca nella comunità scientifica di riferimento e di fornire ai decisori quell'informazione aggiuntiva che serve nelle comparazioni. Ciò, a livello macro, trova conferma in alcuni studi⁴, ma richiede uno sforzo aggiuntivo mano a mano che l'oggetto di osservazione diventa più specifico. Ovvero criteri generali sono utilizzabili a una scala macro ma richiedono puntualizzazioni, e correttivi, a scale inferiori.

Infatti, se avviciniamo la lente vediamo che le diverse discipline non sono affatto così uniformi, che anche all'interno delle scienze dure ci sono aree con comportamenti molto diversi (si pensi ad esempio alle storie: storia della medicina, storia e didattica della fisica, storia della matematica. Oppure a medicina legale, molto più simile come scelta dei canali di comunicazione della ricerca alla giurisprudenza che alla medicina. Analogamente fra le scienze sociali ed umane si trovano settori che somigliano molto alle scienze dure (una parte di psicologia, statistica, una parte di economia, filosofia della scienza, linguistica ecc.).

Un'unica taglia per tutti rischia quindi di fare torto sia all'integrazione tra le “tradizionali” contrapposizioni di “umanisti” e “scienziati”, sia anche agli sforzi che devono essere messi in campo per meglio comprendere le modalità di produzione, diffusione e valorizzazione della ricerca e per combattere gli allineamenti opportunistici e le tendenze omologative che derivano dall'individuazione predefinita di criteri standard.

L'approccio metodologico: il contributo dell'ANVUR

Anche nell'affrontare il tema della valutazione della ricerca nelle scienze umane e sociali si registra un sostanziale ritardo del dibattito nazionale rispetto alla problematica per come è stata affrontata in altri

³ Fonte OECD Science, technology and R&D statistics 2012

⁴ Tra i tanti: Abramo, D'Angelo (2011), VQR 2004-2010, report di area 03,05,06

contesti⁵. Fino al recente documento di Andrea Bonaccorsi⁶ che cerca di affrontare il tema in maniera organica, discutendone i pro e i contro e soprattutto l'opportunità e le "potenzialità e i limiti dell'analisi bibliometrica per le scienze umane"⁷ il dibattito era stato più sui principi e sull'esaltazione delle differenze e i relativi distinguo. Infatti fino alla VQR l'assunto era "non è che le scienze sociali ed umane non sono valutabili con la bibliometria, il problema è che mancano i dati", ovvero la non rappresentatività dell'universo da valutare rispetto ai contenuti delle basi dati utilizzate. Ora che le valutazioni della VQR sono disponibili (in particolare i rapporti di area) appare chiaro che non si tratta solo di disponibilità dei dati, ma di modalità e significato diverso delle citazioni e del loro tempo di maturazione. Nelle scienze umane in generale ma soprattutto nelle singole discipline esse non rappresentano solo un debito intellettuale o la ricerca di conferma delle proprie affermazioni, ma risultano spesso essere citazioni rituali, bibliografiche, oppostive. Si tratta inoltre di mission e target diversi e di comportamenti editoriali multiformi. Non ci si può dunque aspettare che gli umanisti siano soddisfatti dalla adozione di metodi e criteri propri delle scienze dure. Si cercherà di vedere come il fondamento del documento ANVUR sia insufficiente a soddisfare la necessità di un modello dedicato alle scienze umane. Queste aree non rifiutano la valutazione, come spesso e volentieri si sente dire, ma sono alla ricerca di un sistema che possa giudicarle in base alle proprie peculiarità⁸, nella consapevolezza che "unsuitable methods can only lead to debatable conclusions" (KNAW 2005)

L'ottica del documento di Andrea Bonaccorsi è ancora una volta quella di chi si occupa di scienze dure. Come ben dimostrato da Diana Hicks (2004,2013) sono quattro le letterature attraverso le quali la ricerca nelle scienze sociali (ma anche umane) trova la sua espressione: gli articoli su riviste internazionali (in percentuale più o meno significativa a seconda dei settori disciplinari), le monografie (monografie di ricerca, ma anche edizioni critiche, cataloghi, edizioni di fonti), gli articoli su riviste nazionali scientifiche e la letteratura divulgativa. Ciò è dovuto al fatto che le discipline umanistiche hanno un'audience molto più stratificata ed ampia che quella dei pari a livello nazionale e internazionale. Le quattro tipologie si sovrappongono solo molto parzialmente per quanto riguarda citazioni date e ricevute⁹, è quindi necessario, per poter avere un quadro completo, considerarle tutte. Nella realtà inoltre il quadro non ha confini così netti e se possibile si complica ulteriormente. In termini generali fino a qualche anno fa la monografia era il punto di arrivo di un progetto di ricerca che non era preceduto da nulla, mentre ora è abitudine (forzata anche dalla necessità di pubblicare subito per avanzamenti di carriera, presentazione di progetti ecc.) segnare le tappe intermedie di avanzamento degli studi con la pubblicazione di articoli o atti di convegno. Anche la frequente interazione fra gli autori ha portato ad una maggiore circolazione delle conoscenze che ha cambiato le modalità di produzione degli umanisti e degli scienziati sociali, ma più di tutto stanno cambiando le forme di comunicazione dei risultati della ricerca. La digitalizzazione dei contenuti, anche in ambito scientifico, sembra poter garantire oltre a una maggior trasparenza anche una maggiore velocità e un più ampio raggio nella disseminazione delle ricerche a costi molto limitati. Sempre di più la comunicazione di risultati e scoperte passa attraverso blog che segnano le diverse fasi di gestazione di un lavoro, anche in maniera collettiva, prima della concretizzazione in articoli, riviste, saggi.

⁵ Mi riferisco principalmente agli studi in atto da tempo presso la Reale Accademia olandese delle arti e delle scienze (KNAW 2011, 2013) e al lavoro fatto in Norvegia sulla valutazione nelle scienze umane (Sivertsen 2010).

⁶ Bonaccorsi (2012)

⁷ Bonaccorsi (2012)

⁸ In questo senso si è mossa la Reale accademia delle scienze e delle arti già dal 2005 con la pubblicazione dello studio Judging research on its merits

⁹ Tendenzialmente gli articoli citano gli articoli e le monografie citano le monografie (Line 1979)

Fra le proposte di ANVUR (Bonaccorsi 2012) per poter incrementare la quantità di dati a disposizione di chi valuta, c'è la promozione delle riviste di fascia A nei database internazionali¹⁰. Da una sommaria verifica in WOS e Scopus (che sicuramente necessita di approfondimento) notiamo che il numero delle citazioni ricevute dalle riviste italiane presenti in Scopus o Web of Science è molto basso. Le citazioni sono spesso autocitazioni dell'autore, o della rivista, o da riviste italiane non censite nei due database bibliometrici, e quelle da riviste internazionali sono pochissime¹¹. Questo, forse, anche per le tematiche trattate. Ovviamente alcune riviste ottengono un numero maggiore di citazioni perché il subject trattato è di interesse più ampio.¹² La Hicks mette in guardia dall'imporre sistemi di valutazione che "indirizzino" gli studiosi verso temi mainstream poiché lo ritiene snaturante rispetto agli interessi nazionali¹³. I fenomeni socio-culturali sono diversi da regione a regione, e la pubblicazione dei risultati della ricerca in poche riviste di importanza internazionale spesso non è il modo più corretto per la loro disseminazione.

L'altro suggerimento che viene dal documento di Andrea Bonaccorsi è quello del rating¹⁴. Vedremo più sotto come il rating delle riviste abbia rappresentato per alcune aree una attività assai problematica, e soprattutto come ci sia una scarsa correlazione fra rating e giudizi dei peer. Risulta invece più sensato l'auspicio che vengano rese pubbliche le procedure di selezione degli articoli¹⁵.

Rispetto al rating Bonaccorsi sostiene: "Mentre la probabilità di trovare un lavoro scadente in una eccellente rivista non è mai pari a zero, tuttavia è minore della probabilità di trovarlo in una rivista di qualità inferiore"¹⁶. Il confronto fra rating delle riviste e percentuale di articoli eccellenti nelle aree 10-14¹⁷ non dà risultati univoci, poiché, ad esempio nel GEV di area 12 le riviste di fascia A pubblicano i 2/3 degli articoli eccellenti, ma anche il 63 % degli articoli accettabili e il 55% di quelli limitati. Percentuali davvero troppo alte per poter considerare le liste di riviste come riferimento robusto e affidabile per la valutazione.

Dopo la premessa che tiene conto della specificità delle scienze umane per argomenti, ambiti, finalità, target, canali di comunicazione, l'intero documento metodologico di Bonaccorsi si concentra sul tema delle riviste. Il che è interessante, anche per l'ottica con cui si affronta il problema, e i pro e i contro dell'analisi bibliometrica, ma non risolve affatto il problema della valutazione della ricerca nelle scienze umane e sociali. Se è vero che i criteri di valutazione non devono condurre a comportamenti adattivi da parte dei membri delle comunità scientifiche, è altrettanto vero che qui l'esperto di valutazione (che fa parte del Direttivo della Agenzia Nazionale) sembra indicare una strada. Vale a dire siccome l'utilizzo

¹⁰ "In questa linea di attività l'ANVUR intende **sostenere la candidatura di un consistente gruppo di riviste italiane alla indicizzazione presso ISI e Scopus**" (Bonaccorsi 2012). Questo approccio riprende quanto già commentato, ovvero che la difficoltà di utilizzo della bibliometria nelle scienze sociali dipende dalla povertà e parzialità di alimentazione dei data base. Senza negare che il problema esiste (in particolare per tutte quelle discipline e subdiscipline che sono caratterizzate da una barriera linguistica nella loro produzione), si vuole qui sottolineare che, ancor di più in queste aree in cui l'appartenenza alla "classe A" è sostanzialmente l'esito di un processo reputazionale, il rischio di omologazione è molto elevato.

¹¹ Quale può essere inoltre la significatività di una media di citazioni per documento nell'ultimo biennio tendenzialmente al di sotto dello 0? (Si veda SCIMAGO per le riviste italiane nelle Arts and Humanities per il 2012 ad es.)

¹² La Hicks fa l'esempio della rivista scientifica spagnola *Annals of tourism research*, rivista impattata che però su base nazionale riscuote poco interesse da parte degli studiosi, e comunque non tratta temi rappresentativi per la sociologia a livello nazionale (hicks 2013)

¹³ Imposing an evaluation system that privileges international citations will force scholars to choose topics that interest foreign academics... (p. 79)

¹⁴ Per i GEV non bibliometrici è stata prevista la stesura di liste di riviste di fascia A

¹⁵ Anche se dopo la VQR e la ASN ormai molte sedi editoriali dichiarano di fare peer review e di avere procedure di selezione severe.

¹⁶ Bonaccorsi (2012)

¹⁷ Questa analisi è stata fatta da tutti i GEV

del sistema già in atto nelle scienze dure è più semplice, allora pieghiamo le caratteristiche della ricerca alla esigenze della valutazione. Creiamo una banca dati per le scienze umane e sociali e spingiamo i ricercatori a produrre articoli in modo che possano essere più facilmente misurabili. La tentazione è forte e certamente la creazione di una banca dati per le scienze umane è utile (se ne parla a livello europeo da molti anni¹⁸), ma abbiamo già argomentato come la valutazione nelle scienze umane e sociali non possa esaurirsi in un tale strumento.

Anche la scelta degli indicatori per l'abilitazione scientifica nazionale¹⁹ ha dato un forte scossone a quella che è la specificità della comunicazione scientifica nelle scienze umane, favorendo in particolare comportamenti adattivi. In molti Settori Concorsuali infatti un articolo pubblicato su rivista di fascia A vale quanto una monografia di ricerca. L'effetto collaterale, a dire il vero già in atto, è che in un'ottica di puro avanzamento di carriera e ottimizzazione degli sforzi, quello che un giovane ricercatore sarà portato a fare d'ora in poi è pubblicare esattamente su quelle riviste di fascia A, dimenticandosi opere di più ampio respiro ma più impegnative e certamente numericamente meno numerose. Resta da capire se questa sia solo una modalità di comunicazione diversa o snaturi completamente la ricerca in questi ambiti.

I rapporti di area dei GEV di ambito umanistico: dati di contesto

I GEV (Gruppi di esperti della valutazione) dell'ANVUR delle aree 10-14, nei loro rapporti di area relativi alla Valutazione della Qualità della ricerca 2004-2010²⁰, non hanno avuto un atteggiamento univoco nei confronti dell'analisi bibliometrica e quasi solo il GEV 13 ne ha sottolineato (nel rapporto e al di fuori) la bontà²¹. Del resto anche nelle scienze dure gli scienziati si dividono rispetto alla significatività dell'analisi bibliometrica e alla sua capacità di cogliere la qualità della ricerca. Molti considerano la conta delle citazioni come funzione di molte variabili accanto all'impatto scientifico, per cui la probabilità di essere citato dipende da diversi fattori non sempre direttamente collegati al reale valore della ricerca²², come ad esempio il prestigio dell'autore, della istituzione di afferenza, le dimensioni dell'area, la adesione a tematiche mainstream, a scuole di pensiero ecc. Inoltre, poiché le norme e convenzioni per le citazioni non sono fissate, non è del tutto chiaro cosa esse misurino²³, per cui è ormai pensiero comune che la citation analysis può essere significativa specificamente per gruppi numerosi e per archi temporali consistenti.

Ma vediamo cosa ci dicono i dati. I report di ANVUR relativi alla VQR 2004-2010 di dati ne forniscono molti, soprattutto forniscono un commento ampio e dettagliato ad una esperienza di valutazione a dir poco titanica che ci dà un quadro aggiornato sullo stato di salute delle aree umanistiche, ma soprattutto sulle loro peculiarità.

Osservando i dati generali sulle 5 aree possiamo dedurre la seguente tabella:

¹⁸ Martin (2010)

¹⁹ ASN 2012 Indicatori e relative mediane

²⁰ VQR 2004-2010, Rapporti finali GEV e ANVUR

²¹ Checchi (2013), Commissione di concorso della ASN per il Settore concorsuale di scienza delle finanze. Criteri di ammissione <http://www.roars.it/online/lettera-aperta-di-un-gruppo-di-docenti-di-scienza-delle-finanze-sullasn/>

²² Si vedano le distorsioni note con il nome di effetto di S. Matteo, in particolare Larivière e Gingras (2010)

²³ Lee, Sugimoto, Zhang e Cronin (2012)

Area	SSD	Prodotti attesi	Prodotti valutati	Articoli	Contributi in volume	Monografie (ediz critiche)	metodo di valutazione	%eccellenti	prodotti in altre lingue ²⁴
10	77	14.637	13966	26,50 %	46,99%	23,61%	PR	46% ediz critiche (23,27% altre tipologie)	17% (inglese) 12% altre lingue
11	34	13487	13152	23,90 %	39,8	33,06	PR (Biblio per PSI e M-EDF)	15% area non biblio 34% area biblio	12 % non biblio inglese 62% biblio inglese
12	21	12798	11882	32,00 %	39%	29%	PR	15% monogr. 11,3 % articoli	5% (inglese)
13	19	12654	11941	62%	23%	13%	PR (non articles) informed PR (articles)	18%	56% (inglese)
14	14	4494	4327	28,60 %	32,60%	33,60%	Informed PR con 6% in biblio	8%	21% (inglese)

Il primo commento che si può fare a questi dati è che nelle aree 10-14 i lavori di ampio respiro hanno (o almeno avevano fino al 2010) una grande rilevanza per gli autori²⁵, che li hanno preferiti a lavori più brevi. La lingua utilizzata è nella maggior parte dei casi l'italiano, in parte perché lingua franca nell'ambito disciplinare, in parte perché spesso i temi trattati sono legati a realtà locali.

Per quanto riguarda l'area 10 il GEV sottolinea alcuni problemi emersi durante l'esercizio. In primo luogo l'inadeguatezza della scheda di valutazione²⁶ con i tre criteri di rilevanza, novità, internazionalizzazione, non sembra particolarmente adatta per l'area 10. In particolare si rileva che non

²⁴ Si tenga conto che il campo lingua non era obbligatorio e in alcuni casi non è stato compilato.

²⁵ Anche se in alcune aree si segnalano una diminuzione delle monografie presentate a mano a mano che ci si avvicina nel tempo

²⁶ Anvur ha predisposto una scheda di valutazione per i peer reviewer. La scheda conteneva tre domande a cui rispondere attribuendo a ciascuno dei punti un punteggio da 0 a 3 e un commento finale facoltativo (per le dimensioni da valutare di veda il Bando VQR 2004-2010 a pagina 7 http://www.anvur.org/attachments/article/122/bando_vqr_def_07_11.pdf

è possibile usare la stessa griglia per tipologie di lavori così diversi fra loro (monografie, articoli, capitoli). Il secondo problema è legato alla verifica della corrispondenza fra valutazione dei pari e rating delle riviste. Il GEV fa presente di aver portato a termine un rating delle riviste in accordo con le società scientifiche, ma di non averne tenuto conto²⁷. In molti casi non c'è corrispondenza fra la valutazione eccellente e la collocazione del lavoro valutato in una rivista di fascia A. Il GEV ravvisa una distribuzione molto differente della produttività fra nord e sud ma sottolinea la presenza di un risultato medio buono.

Il rapporto per l'area 11 individua un problema nel fatto che la peer review, applicata in maniera così estensiva, non garantisce l'uniformità dei giudizi che sono sempre ovviamente soggettivi, e rileva le stesse difficoltà dell'area 10 con una scheda di valutazione troppo generica, quindi suscettibile di interpretazioni assai diverse che ha condotto a giudizi molto difformi. Per i SSD molto piccoli si sottolinea la tendenza ad avere una concentrazione di giudizi eccellenti. In generale la peer review attribuisce giudizi più severi dell'analisi bibliometrica e concentra i voti nelle fasce centrali (buono e accettabile) piuttosto che in quelle estreme (eccellente e limitato). Il GEV esprime una grande soddisfazione per il rating delle riviste che distingue riviste nazionali e internazionali, settoriali e intersettoriali di fascia A e B per ciascuno dei quattro subGEV in cui il GEV si è suddiviso²⁸. Esiste una buona correlazione tra i lavori valutati eccellenti dalla peer review e la loro presenza sulle riviste internazionali di fascia A, mentre le riviste di fascia B vengono equiparate alle riviste scientifiche. I lavori pubblicati su quelle che sono state definite riviste internazionali di fascia A hanno poi un tasso di prodotti eccellenti quasi del 50%, mentre la categoria B+C registra solo un 10% di lavori eccellenti.

Al GEV 11 sono stati anche sottoposti i lavori di alcune storie (storia della medicina, della matematica, storia e didattica della fisica) che, facendo parte di aree bibliometriche non avevano un proprio rating per le riviste di riferimento. Su queste si è lavorato solo in peer review, e resta dunque il duplice problema di prevedere per il futuro delle liste anche per queste discipline e soprattutto di pensare ad una revisione della struttura delle diverse aree. Per quanto riguarda invece la coerenza dei giudizi dei pari fra di loro, il diverso peso attribuito dai revisori ai tre parametri previsti dalla griglia di valutazione ha condotto a esiti molto difformi che è stato necessario ricomporre con l'aiuto di consensus group.

Sempre nel GEV 11 le discipline di Psicologia e Metodi e didattica delle attività motorie hanno avuto una valutazione bibliometrica rispettivamente per il 59% e il 78% che non sono affatto la totalità dei prodotti, ciò ad indicare ancora il forte radicamento a modalità di pubblicazione e comunicazione dei risultati tipiche delle aree umanistiche.

Nonostante la grande enfasi posta sul rating delle riviste, anche in questa area i dati ci dicono che gli articoli presentati sono addirittura meno che nell'area 10, le monografie sono in numero maggiore i contributi in volume in numero leggermente inferiore. Le monografie restano dunque la tipologia più significativa anche per questa area e risulta quindi evidente la insufficienza di un sistema di rating delle riviste per la valutazione dell'area.

Sul criterio di internazionalizzazione presente nella griglia, anche in questo caso ci sono state interpretazioni diverse che hanno pesato positivamente o negativamente sul giudizio finale e quindi si

²⁷ Come è stato dichiarato ad apertura del documento che, sul sito ANVUR VQR, ne riporta i risultati, la classificazione risultante non è stata considerata uno strumento di valutazione della produzione scientifica del periodo 2004-2010 (cfr. Criteri di Valutazione del GEV 10, al punto 3.3)

²⁸ Antropologia Geografia e Storia, Filosofia, Pedagogia, Psicologia

sottolinea la necessità di una più chiara definizione di cosa significhi questo concetto per l'area. Infine comune ad area 10 e 11 è l'attribuzione di voti decisamente più alti nei settori molto piccoli. Un fenomeno riconosciuto anche a livello internazionale.

Il rapporto sottolinea come l'analisi puntuale dei risultati delle valutazioni debba essere oggetto di studio di altri, tuttavia in parecchi punti, anche citando pareri di soggetti valutati²⁹ sembra suggerire quale debba essere la via maestra. Le monografie vengono per la maggior parte definite (dal rapporto) come pubblicate a fini concorsuali e i contributi in volume sono volumi congressuali, d'occasione, dove si scrive su invito, e legati all'appartenenza a determinate reti. Il quadro che emerge da quella che dovrebbe essere una analisi oggettiva che lascia ad altri l'interpretazione dei dati è quello di una serie di comunità scientifiche isolate le une dalle altre e dalla comunità internazionale, che pubblicano tendenzialmente per motivi di carriera in sedi poco appropriate e spesso a pagamento. L'alternativa a questo tipo di atteggiamento sembra essere la pubblicazione in riviste indicizzate e internazionali (le riviste di fascia A internazionali). Ma non tutti i tipi di ricerche sono comprimibili in questa tipologia di lavori. Forse la pubblicazione ad accesso aperto di volumi, saggi o articoli potrebbe portare ad un maggiore livello di diffusione al di là dei confini nazionali. A ben vedere forse le modalità di pubblicazione spesso ancora su carta sono un ostacolo, maggiore anche di quello linguistico, ad una ampia disseminazione dei risultati in queste aree. Con una modalità di pubblicazione ad accesso aperto i lavori sarebbero letti da più persone, qui e all'estero, senza bisogno di sottoscrizione, e quindi citate criticate ecc. molto di più che attraverso la pubblicazione in riviste internazionali che poi solo pochi atenei potrebbero permettersi di acquistare.

Anche nelle conclusioni il GEV sottolinea che la peer review ha grosse bias, difficilmente sanabili. Le alternative proposte sono quelle di rivolgersi alle banche dati internazionali o eventualmente alla costituzione di una banca dati nazionale che registri le citazioni anche delle riviste italiane. Questo tipo di proposte, alla luce dei dati messi a disposizione, appaiono poco aderenti alla realtà. Il GEV non pare in grado di uscire dagli schemi proposti per le scienze dure, e fornisce ai giovani studiosi un modello che snatura la ricerca di ambito umanistico, o meglio che si adatta ad alcuni settori, ma non rispecchia minimamente le caratteristiche di altri.

Il GEV 12 ha dichiarato la sua apertura ad accogliere pubblicazioni (ad esempio i manuali) non del tutto citate nel bando, ma neppure espressamente vietate, sottolineando l'importanza fondamentale delle monografie per le discipline giuridiche. Anche questo GEV si è affidato alla peer review, sottolineando come l'attività di rating delle riviste sia stata una attività aggiuntiva rispetto alla valutazione qualitativa, e del tutto sperimentale³⁰. L'attività di classificazione delle riviste è stata intesa come gesto dovuto rispetto a quanto previsto dal bando, e con l'intento di fornire ulteriori elementi conoscitivi, non vincolanti, a chi doveva valutare i lavori. Anche il GEV 12 segnala la difficoltà di utilizzo della griglia di valutazione ed in particolare la definizione del concetto di internazionalizzazione.

²⁹ “Non ho cioè tenuto conto del fatto che uno dei doveri di chi si impegna scientificamente è non solo quello di produrre personalmente buoni risultati scientifici, ma anche di far progredire la disciplina nel suo insieme, diffondendo il più possibile tali risultati: ciò che si ottiene, oltre che con le monografie, cercando di pubblicare su riviste di prestigio, abitualmente frequentate anche da colleghi stranieri “(rapporto di area 11 p.63).

³⁰ È opinione diffusa, tanto nei Paesi di civil law, quanto in quelli di common law (con la parziale eccezione degli Stati Uniti) che il giudizio di merito riferito alla “qualità” della sede in quella quale un determinato lavoro scientifico è pubblicato non sia direttamente e completamente trasferibile a quel lavoro (GEV di area 12 rapporto finale p. 21)

L'area 13 dichiara il proprio entusiasmo nella attività di redazione di una classificazione di riviste e nella adozione della stessa per la valutazione degli articoli³¹, rilevando poi una buona correlazione fra esiti della peer review e valutazione bibliometrica dei lavori presentati. Sono stati valutati con peer review tutti i lavori non pubblicati su riviste (o sulle riviste comprese nella lista predisposta). Il GEV rileva con soddisfazione il fatto che vi sia un buon numero di articoli sottoposti che è cresciuto in quasi tutti i settori dal 2004 al 2010 mentre è diminuito il numero delle monografie e vede questo come il segno di un avvicinamento alla modalità di comunicazione delle scienze dure. In realtà bisogna vedere anche quanto la stesura di articoli sia legata ai sistemi di valutazione interni³²

Anche il GEV 14 rileva la diminuzione delle monografie presentate nel settennio. Per gli anni più recenti i ricercatori hanno scelto di presentare articoli³³. Con l'aumento del numero degli articoli si segnala l'aumento dei lavori di lingua inglese. La lista di riviste predisposta dal GEV ha suscitato non poche polemiche³⁴ sia rispetto alla considerazione del parere delle società scientifiche, sia rispetto alla modalità di redazione della lista stessa. Il GEV ha svolto le valutazioni secondo il sistema della informed peer review, verificando poi, come per tutti gli altri GEV, la corrispondenza fra giudizi eccellenti e collocazione dell'articolo nella sede editoriale. Il 36,6% degli articoli eccellenti è stato pubblicato in riviste presenti nel rating del GEV (ABC) ma solo il 6,8% degli articoli pubblicati su rivista di fascia A è stato reputato eccellente. Con dati così bassi non si può assolutamente pensare di svincolarsi dalla peer review. Tale giudizio è stato confermato anche dal confronto fra risultati della peer review e analisi bibliometrica per la piccola parte destinata al confronto fra le due modalità di valutazione.

In generale si può dire che l'enfasi posta sulla classificazione delle riviste, in particolare da alcune aree, ha fatto sì che molti studiosi che fino ad ora avevano disdegnato il genere dell'articolo, improvvisamente riscoprissero la loro vocazione alla stesura di questo genere accademicamente più redditizio, creando comportamenti adattivi. In questo sono stati molto aiutati da alcune commissioni di concorso o da alcuni atenei che ritengono ad esempio come fondamentale il superamento di tutte e tre gli indicatori per la abilitazione scientifica nazionale.

Più ancora che per i numeri e le classifiche, i rapporti della VQR sono utili per delineare una realtà delle singole aree estremamente variegata, in particolare per le scienze umane e sociali (che abbiamo per questo motivo analizzato separatamente), un quadro che risulta difficilmente comprimibile negli schemi e nelle metodologie adottate ormai da anni per le scienze dure e che deve necessariamente spingere ad una riflessione. Laddove l'agenzia nazionale ha prodotto come unico documento metodologico uno studio sull'impiego dei metodi bibliometrici nelle scienze umane, altrove si è da tempo capita l'importanza di queste aree per lo sviluppo del pensiero e della cultura e il loro carattere multiforme, e si è proceduto ad analisi molto più raffinate. Vale la pena di citare qui il modello sviluppato dalla Reale accademia di scienze e lettere per il sistema della ricerca olandese (KNAW 2011), un modello che ha avuto una lunga gestazione, che ha visto la somministrazione di questionari a livello internazionale e di

³¹ Si veda il documento GEV 13, Classification Imputation and ranking of Journals (sept. 2012)

³² In alcuni atenei per ottenere finanziamenti è necessario pubblicare articoli su riviste incluse in liste predisposte

³³ Va ricordato come una delle caratteristiche della produzione di risultati in ambito umanistico sia quella di raccogliere, razionalizzare e sistematizzare i contributi precedentemente elaborati e farli confluire in successive opere monografiche. Da qui anche la prevalenza di più recenti articoli che potrebbero poi essere compresi in monografie.

³⁴ Si veda la lettera di alcuni direttori responsabili pubblicata su voce sociologica

<http://www.vocesociologica.it/notizie/133-gev14.html> e i commenti di Maria Chiara Pievatolo

<http://btfp.sp.unipi.it/?p=1938>

interviste alle comunità di riferimento olandesi, alla ricerca di una definizione di qualità. Qualità della ricerca ma anche qualità sociale (ponendo l'accento sul tema della valorizzazione della conoscenza).

In effetti esiste una qualità intrinseca che solo un pari può cogliere (ovviamente se vuole coglierla), ma anche una qualità esteriore che è legata alla sede di pubblicazione, o ai premi e riconoscimenti ricevuti.

La peer review resta il sistema portante del modello di valutazione nelle scienze umane, ma se ne riconoscono i limiti. Alcuni sono determinati da situazioni contingenti: la sollecitazione a pubblicare sempre di più ha portato ad un aumento globale della produzione scientifica che deve essere valutata dai peer e che però sono anch'essi ricercatori e soggetti dunque alla stessa urgenza. Il sistema rischia così di implodere con la possibilità anche di una diminuzione della qualità dei lavori pubblicati e del controllo da parte della comunità scientifica su ciò che circola. La peer review presenta inoltre bias legate alla soggettività di chi giudica, e c'è quindi necessità di una maggiore standardizzazione nelle pratiche³⁵. Un aiuto può venire dagli indicatori qualitativi, che gli olandesi individuano in criteri legati ai risultati scientifici distinti in output indicators (recensioni, pubblicazioni in riviste di fascia A, citazioni), indicatori di prestigio (keynote, presenza in board prestigiosi), indicatori d'uso (citazioni, reviews ecc.³⁶), e criteri legati all'impatto sociale distinti in output per un ampio pubblico, rapporti con ed effetti sulla società civile, riconoscimento sociale. La bibliometria può fornire informazioni utili, ma rappresenta uno degli elementi che compongono un quadro complesso, insieme ad esempio alla rilevazione dell'impatto sociale, individuabile come contributo al dibattito politico, culturale e al dibattito sull'identità di una nazione.

Le liste di sedi editoriali e il loro rating (in A, B e C) possono anch'esse dare un contributo, se elaborate in maniera scientifica, condivisa e con la disponibilità ad una revisione continua e se utilizzate conoscendone i limiti (il giudizio viene dato sul contenitore, non su ogni contenuto) a supporto per la definizione delle caratteristiche della produttività di un ricercatore (avere parecchi articoli in riviste di fascia A può essere un segno di qualità della ricerca svolta più che averne solo uno)³⁷.

Non c'è motivo per bandire l'una o l'altra famiglia di indicatori (peer review, bibliometria, liste di riviste ed editori, impatto sociale). Quello che si richiede è la disponibilità di un ampio set di indicatori entro i quali la scelta (o la compatibilità/comparabilità) va guidata da un processo che associ gli obiettivi da perseguire nel percorso valutativo alla qualità e significatività degli strumenti utilizzati. Ciò deve potersi tradurre in una specifica attenzione alla disciplina, al contesto e alle altre dimensioni valutative attraverso un mix flessibile e rivedibile con opportuna frequenza.

Un altro elemento da tenere in conto nel disegno di un modello di valutazione per le scienze umane è la necessità di attribuire il peso corretto alle diverse tipologie di lavori. Un articolo non può avere lo stesso peso di una monografia. Una monografia di ricerca dovrebbe avere un peso diverso da una edizione critica. Una monografia pubblicata presso un editore che pubblica abitualmente autori di varie nazionalità ha un peso diverso da una monografia pubblicata presso un editore che pubblica autori italiani.³⁸

Sulla base di quanto riportato dai GEV nelle diverse aree, la valutazione nelle scienze umane è ancora alla ricerca di una visione unitaria e rappresentativa. Gli strumenti, i metodi e gli indicatori sono tutti da

³⁵ In questo senso in molti sistemi di valutazione è richiesta anche la motivazione del giudizio assegnato

³⁶ KNAW (2011)

³⁷ Sull'opportunità che queste liste vengano gestite a livello nazionale piuttosto che internazionale vedi Knaw 2011. P. 40-41

³⁸ Si veda il sistema sviluppato in Norvegia (Sivertsen 2010)

costruire e, soprattutto, da condividere, tenendo in conto il possibile contributo delle comunità scientifiche disciplinari. Ci sono dei modelli a cui ispirarsi, elaborati da paesi in cui la riflessione è iniziata da parecchi anni e che quindi potrebbero essere adattati alla realtà italiana, ma ciò può essere fatto solo se si riconosce alle scienze umane e sociali la propria peculiarità.

Si aprono quindi opzioni che meritano alcune riflessioni.

In primo luogo va definito il percorso che porti a un miglioramento (copertura, validazione, significatività e utilizzo) delle informazioni relative alle pubblicazioni/output e al loro inserimento in database. La non definita rilevanza “commerciale” di questi dati penalizza la motivazione e l’interessamento dei gestori dei db esistenti mentre la forte polarizzazione nazionale ostacola un percorso armonizzato a livello europeo. Ma anche venisse finalmente avviato e realizzato il processo, per le considerazioni precedentemente presentate, la “bibliometria tradizionale” anche adattata poco si presta ad analisi comparative. Rimane quindi il tema dell’individuazione di indicatori efficaci a livello di subarea, non omologabili in un astratto quadro comune ma interpretabili per confronti.

L’esperienza realizzata col rating delle riviste ai fini VQR e per mezzo delle società scientifiche non rappresenta certamente un modello esente da critiche, sia nel metodo che nel merito delle proposte. Ma, inevitabilmente, riflette il diverso grado di maturazione, e di credibilità su questi temi, che ognuno dei soggetti ha deciso di evidenziare. Il terreno intermedio tra peer review e bibliometria è quanto mai significativo per queste aree e una sistematica mappatura dei metodi e dei problemi da affrontare potrebbe portare alla definizione di linee guida comportamentali che potrebbero trovare, nella valutazione interna di grandi atenei, o grandi enti di ricerca, generalisti un terreno concreto di sperimentazione. L’esperienza olandese suggerisce però che solo la dimensione di sistema è quella che consente di far fare un effettivo passo in avanti a tutti. E a questo bisogna puntare, magari traendo spunto dalla pluralità di esperienze e competenze metodologiche sviluppate all’interno di associazioni scientifiche³⁹.

Bibliografia

Abramo Giovanni, D’Angelo Ciriaco Andrea (2011) Evaluating research: from informed peer review to bibliometrics, *Scientometrics*, 3(87) pp 499-514

ASN (2012) indicatori e relative mediane, recuperabile all’indirizzo:

http://www.anvur.org/index.php?option=com_content&view=article&id=253&Itemid=314&lang=it

Bonaccorsi Andrea (2012) Potenzialità e limiti dell’analisi bibliometrica nelle aree umanistiche e sociali. Verso un programma di Lavoro. ANVUR

Cecchi Daniele (2013) “Valutazione: c’è un medicinale adatto?” *Il Mulino* 2(2013) Doi: 10.1402/72992

Hicks Diana (2004), ‘The four Literatures of Social sciences’ , in Moad Henk ed) *Handbook of Quantitative Science and technology research*, Kluwer academic

³⁹ Una versione molto abbreviata di questo articolo è stata presentata al Workshop organizzato da AIS AIV “Criteri, indicatori e strumenti per la valutazione della qualità della didattica e della ricerca nell’Università”, Firenze 11 ottobre 2013

Hicks Diana (2013), "One size doesn't fit all: on the co-evolution of national evaluation systems and social science publishing", *Confero*, 1, 1 pp. 67-90

Larivière Vincent e Gingras Yves (2010), The impact factor's Matthew Effect: A natural experiment in bibliometrics, *JASIST* 2(61), pp.424-427

Lee Carole J., Sugimoto Cassidy R., Zhang Guo, Cronin Blaise, Bias in peer review, *JASIST*, 1 (64), pp. 2-17

Line, Maurice B. (1979), "The influence of the type of sources used in the results of citation analyses" *The journal of documentation*, 4, 35, pp 265-284.

KNAW (2005) Judging research on its merits recuperabile all'indirizzo: <https://www.know.nl/shared/resources/actueel/publicaties/pdf/20051029.pdf>

KNAW (2011) Quality indicators for research in the Humanities, recuperabile all'indirizzo: <https://www.know.nl/en/actueel/publicaties/quality-indicators-for-research-in-the-humanities>

KNAW (2013) Towards a framework for the quality assessment of the social sciences, recuperabile all'indirizzo: <https://www.know.nl/en/actueel/publicaties/towards-a-framework-for-the-quality-assessment-of-social-science-research>

Martin Ben (2010) Towards a bibliometric database for social sciences and the humanities, a European scoping project.

Sivertsen Gunnar (2010), A performance indicator based on complete data for the scientific publication output at research institutions, *ISSI newsletter*, 1 (6), pp. 22-28

VQR 2004-2010 (2013), Rapporti finali GEV e ANVUR, recuperabile all'indirizzo: http://www.anvur.org/index.php?option=com_content&view=article&id=245&Itemid=198&lang=it