

La peer review nella valutazione della ricerca e data analysis sulle pubblicazioni del SSD M-STO/08
(Archivistica, bibliografia e biblioteconomia)

Original

La peer review nella valutazione della ricerca e data analysis sulle pubblicazioni del SSD M-STO/08 (Archivistica, bibliografia e biblioteconomia) / Morriello, Rossana. - In: JLIS.IT. - ISSN 2038-1026. - 14:1(2023), pp. 99-120. [10.36253/jlis.it-510]

Availability:

This version is available at: 11583/2975102 since: 2023-01-24T18:06:13Z

Publisher:

FUP

Published

DOI:10.36253/jlis.it-510

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

Peer review in research assessment and data analysis of Italian publications in SSD M-STO/08 (Archival science, bibliography, library science)

Rossana Morriello^(a)

a) Politecnico di Torino, <https://orcid.org/0000-0002-9990-9243>

Contact: Rossana Morriello, rossana.morriello@polito.it

Received: 1 October 2022; Accepted: 26 October 2022; First Published: 15 January 2023

ABSTRACT

Since the introduction of research assessment systems at institutional level in the 1980s, the ongoing debate on the roles and functions of peer review and bibliometrics has been vivid and lively. In the first part of the article, the main lines over time of this debate are traced, and a reflection on the epistemic functions of peer review and citations is proposed. In Italy, the first research assessment exercise (VTR) was based on peer review only, while the following ones (VQR) were based on different methods for bibliometric disciplines and non-bibliometric disciplines, namely bibliometric indicators and peer review. Starting from a data analysis on Italian publications, and using as a sample data from M-STO/08 (Archival science, bibliography and library science) area, the essay shows some trends and changes in publication habits in HSS. Conclusions open a perspective on revitalization of peer review as a solid qualitative method for research assessment.

KEYWORDS

Peer review; Bibliometrics; Research assessment; VQR; Academic field M-STO/08 (Archival science, bibliography and library science); HSS; Humanities and social sciences; Academic research.

La peer review nella valutazione della ricerca e data analysis sulle pubblicazioni del SSD M-STO/08 (Archivistica, bibliografia e biblioteconomia)

ABSTRACT

Fin dall'introduzione dei sistemi di valutazione della ricerca a livello istituzionale negli anni Ottanta del XX secolo, è stato vivo e acceso il dibattito sul ruolo della peer review rispetto agli indicatori bibliometrici, di cui nella prima parte del saggio si riportano le linee principali, affiancate da una riflessione sulle funzioni epistemiche della peer review e della citazione. In Italia, il primo esercizio di Valutazione Triennale della Ricerca (VTR) era basato esclusivamente sulla peer review mentre nei successivi esercizi di Valutazione della Qualità della Ricerca (VQR) si è consolidata una distinzione tra gli indicatori per i settori bibliometrici e la peer review per i settori non bibliometrici. Partendo dai dati dall'analisi dei dati sulle pubblicazioni italiane negli anni della VQR, esemplificate tramite l'SSD M-STO/08 (Archivistica, bibliografia e biblioteconomia), il saggio si sofferma poi su alcune tendenze e cambiamenti rilevabili nelle abitudini di pubblicazione dei settori HSS. Le conclusioni aprono una prospettiva di rivalutazione della peer review come solido strumento qualitativo per la valutazione della ricerca.

PAROLE CHIAVE

Peer review; Bibliometria; Valutazione della ricerca; VQR; Settore scientifico M-STO/08 (Archivistica, bibliografia e biblioteconomia); HSS; Scienze sociali e umanistiche; Ricerca accademica.

Alle origini della valutazione della ricerca

Se volessimo individuare un punto di svolta per la valutazione della ricerca lo potremmo collocare nel secondo dopoguerra quando, con il passaggio dalla piccola scienza alla grande scienza teorizzato da Derek de Solla Price, si registra un cambiamento nelle forme organizzative e nelle pratiche della ricerca scientifica, soprattutto nei settori STM (Science Technology Medicine). La ricerca della “grande scienza” viene condotta tramite reti di collaborazione estese, anche internazionali, e non più individualmente dal singolo ricercatore, ed è condizionata dalle esigenze di tipo politico e socio-economico, discostandosi così dal fondamento della libertà di ricerca del singolo ricercatore (Price 1967; Greco 2013 e 2019; Villa 2018). La guerra, e in particolare il Progetto Manhattan, aveva, infatti, generato una rinnovata consapevolezza del ruolo cruciale della ricerca scientifica nella società, e quindi della valenza economica e politica, soprattutto della ricerca applicata. Tale consapevolezza indusse i governi a investire maggiormente nella ricerca e nelle università (in misura e con modalità diverse nei diversi paesi), e il numero di docenti e ricercatori cominciò a crescere. Di conseguenza, aumentò anche il numero delle pubblicazioni, e a sua volta il numero delle riviste scientifiche poiché gli editori colsero e assecondarono l’esigenza di aprire nuove sedi editoriali per accogliere i risultati della ricerca. Difatti, sulla base dei calcoli di Price, in quegli anni «il numero degli scienziati era passato da 50.000 alla fine del XIX secolo a più di un milione; il numero delle riviste scientifiche era passato da circa 100 nel 1830 a diverse decine di migliaia» (Faggiolani 2015, 32).

Negli Stati Uniti, dove gli investimenti sulla ricerca furono consistenti anche grazie alla politica mirata di alcune figure di riferimento come Vannevar Bush (Bush 2013), si iniziò a studiare sistemi in grado di velocizzare e rendere oggettiva la valutazione di quantità crescenti di progetti e di pubblicazioni sottoposte alle istituzioni per richiedere finanziamenti. Fin dagli anni Venti del XX secolo, diversi studiosi avevano analizzato le possibilità di usare indicatori quantitativi derivati dalle citazioni per determinare l’utilità di una pubblicazione (Gingras 2016), ma nel dopoguerra la necessità di far fronte al consistente aumento dei prodotti della ricerca ampliò il dibattito teorico e le sperimentazioni, e pose le basi per lo sviluppo della bibliometria¹. Leo Meltzer, sociologo e psicologo alla Cornell University, fu tra i primi a parlare di valutazione basata sulle citazioni. Nel suo articolo del 1956, *Scientific Productivity in Organizational Settings* (Meltzer 1956), egli individuava come indicatori significativi la quantità di articoli e la quantità di citazioni ricevute, inserendosi nel dibattito emergente in quegli anni sull’efficacia delle grandi reti di collaborazione rispetto alla scienza condotta individualmente². Egli assumeva la citazione come un indicatore di qualità e desumeva che anche il numero di pubblicazioni fosse correlato alla qualità dei lavori prodotti da uno scienziato. Contemporaneamente, Eugene Garfield stava collaborando alla sistematizzazione

¹ Non è casuale, infatti, la nascita nel 1969, del termine bibliometria coniato da Alan Pritchard con il suffisso greco a rimarcare l’aspetto di misurazione. Nel suo articolo su *Journal of Documentation*, Pritchard propose di sostituire l’espressione *statistical bibliography*, coniata nel 1922 da E. Wyndham Hulme e precedentemente in uso, con il nuovo termine *bibliometrics* (Pritchard 1969).

² Secondo Meltzer, la forma organizzativa non incideva direttamente sulla produttività scientifica dei ricercatori ma agiva rispetto a due variabili ugualmente impattanti e inversamente proporzionali: i finanziamenti e la libertà. Nelle grandi reti collaborative vi sono più fondi e attrezzature disponibili ma minore libertà rispetto ai temi di ricerca e alla creatività scientifica, mentre lo scienziato che lavora singolarmente dispone di meno fondi ma ha maggiore libertà.

di strumenti di recupero veloce delle informazioni su autori e articoli nell'ambito del progetto Welch Medical Library Indexing della Armed Forces Medical Library (che poi diverrà la National Library of Medicine) e ponendo le basi per la nascita dello *Science Citation Index* e dei successivi altri indici citazionali³.

Tuttavia, negli anni Sessanta e Settanta del XX secolo, la valutazione della ricerca consisteva prevalentemente in processi indirizzati ai soli ricercatori e gruppi di ricerca che sottoponevano progetti per ottenere finanziamenti pubblici, e gli strumenti sviluppati venivano usati prevalentemente a tale scopo. Dagli anni Ottanta si cominciò, invece, a parlare di valutazione istituzionale dei dipartimenti e degli atenei, come sommatoria delle valutazioni individuali, e condotta dalle agenzie governative per misurare l'efficacia e l'efficienza delle strutture di ricerca. In particolare, il primo esercizio di valutazione di questo tipo fu introdotto nel 1986 dal governo Thatcher in Gran Bretagna. Gli anni Ottanta furono un'epoca di mutamenti socio-politici significativi, nei quali l'idea di collettività veniva messa in ombra dall'individualismo crescente e le ineguaglianze risultavano esaltate dall'azione governativa, fino a divenire quasi un valore. In Gran Bretagna, furono anche gli anni della forte spinta verso le privatizzazioni, cui si affiancava un rafforzamento del potere del governo centrale rispetto alle istituzioni locali. La valutazione della ricerca andava a inserirsi in questo contesto utilitaristico, dominato dall'ossessione sulla misurazione dell'efficacia, dell'efficienza e della produttività attraverso indicatori di performance. In tale contesto gli indicatori bibliometrici rappresentavano lo strumento ideale, e soprattutto già disponibile.

Una ricerca dell'espressione "research assessment" su Google Books Ngram Viewer mostra, infatti, l'impennata dell'uso del termine proprio tra gli anni Settanta e gli anni Ottanta del XX secolo (figura 1.1).

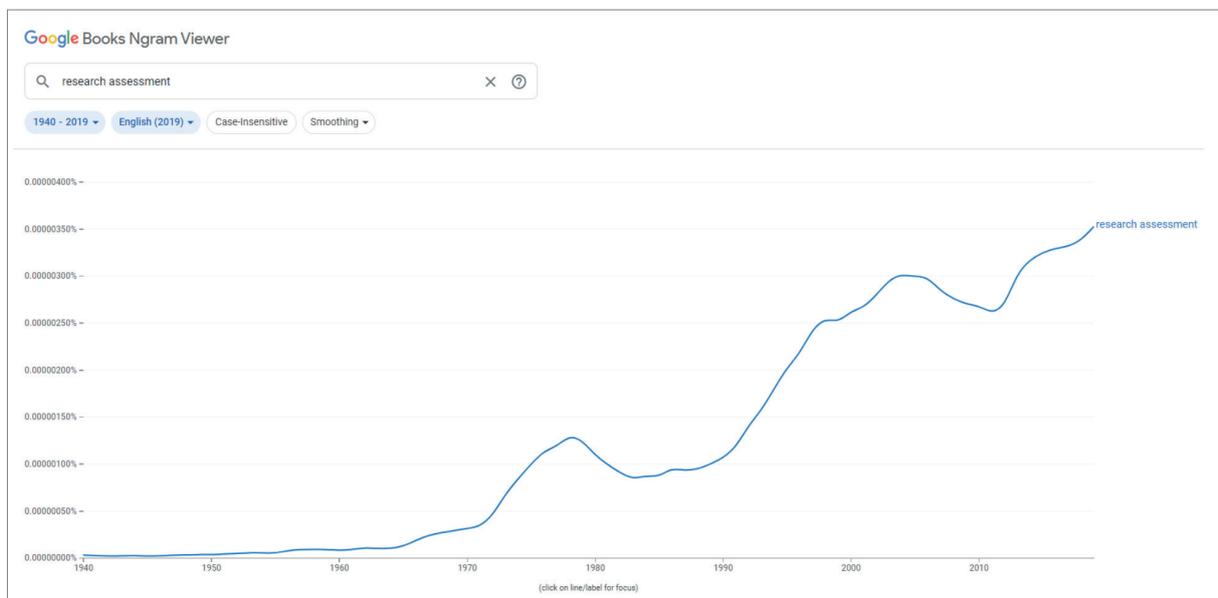


Figura 1. Ricerca della stringa "research assessment" in Google Ngram Viewer

³ Le opere di Garfield sono raccolte nel sito a lui dedicato <http://www.garfield.library.upenn.edu/>.

La Gran Bretagna fu dunque il primo paese europeo a introdurre in forma istituzionalizzata la valutazione della ricerca e a collegarla ai finanziamenti con il primo esercizio di valutazione triennale, il Research Assessment Exercise (RAE),⁴ che combinava peer review e indicatori quantitativi. Questo sollevò subito una serie di dubbi e di contestazioni, sia in relazione alla poca trasparenza dei criteri di valutazione, sia per le note criticità degli indicatori bibliometrici, e infine rispetto ai risultati penalizzanti per molti atenei in termini di allocazione dei fondi (Phillimore 1989; Bence e Oppenheim 2005; Weingart 2005). Una delle conseguenze fortemente criticate riguardò la progressiva mutazione nei principi alla base della gestione dell'università verso un approccio sempre più manageriale, con l'incremento della concorrenza e della competizione per i finanziamenti. Un approccio che finiva col porre al centro dell'attenzione l'istituzione rispetto ai singoli e alle discipline, andando a definire la valutazione istituzionale come sommatoria delle valutazioni individuali dei ricercatori. In questa prospettiva, le differenze esistenti tra le discipline rispetto alle modalità di produzione e circolazione della conoscenza, alle abitudini citazionali, alle pratiche di conduzione della ricerca, potevano essere livellate e non prese in considerazione, poiché non funzionali all'obiettivo. Gli strumenti quantitativi assolvevano efficacemente il compito di misurazione, ma al contempo suscitavano delle perplessità nella comunità scientifica. Con l'introduzione del primo esercizio di valutazione delle istituzioni della ricerca si accentuò, infatti, anche il dibattito sull'opportunità e sull'effettiva efficacia della bibliometria rispetto al metodo tradizionale della peer review, che avrà influenza anche sullo sviluppo del sistema di valutazione della ricerca italiano.

La valutazione della ricerca in Italia: dalla VTR alla VQR

In Italia, il primo esercizio di valutazione della ricerca nazionale è stato la VTR (Valutazione Triennale della Ricerca), condotto dal Comitato di Indirizzo per la Valutazione della Ricerca (CIVR) istituito presso il Ministero dell'Istruzione, dell'Università e della Ricerca dal Decreto Legislativo 104 del 1998, e successive modifiche, e normato nel suo funzionamento dal decreto ministeriale del 24 novembre 1999. L'esercizio di valutazione si ispirava a quanto realizzato in Gran Bretagna, ma nasceva anche come conseguenza dell'introduzione di un nuovo sistema di finanziamento da parte del Ministero nel 1993, denominato Fondo di Finanziamento Ordinario (FFO). Il nuovo sistema affiancava alla ripartizione dei fondi ministeriali sulla base dei dati storici una quota allocata in base alla misurazione dei risultati della ricerca. Tale quota, inizialmente minima, fu aumentata progressivamente a partire dal 2008 da varie leggi (Geuna e Piolatto 2016). Gli esercizi di valutazione risultano quindi fin dall'inizio collegati alla ripartizione dei finanziamenti. Questa è una delle criticità del nostro sistema di valutazione e in generale dei *Performance-based Research Funding Systems*, poiché spostano l'obiettivo della valutazione e ne deformano la natura. Difatti, la maggior parte dei sistemi nazionali di valutazione della ricerca è svincolato dai finanziamenti e assolve funzioni diverse, quali rendicontare pubblicamente l'attività di ricerca per giustificare nei confronti dei cittadini gli investimenti pubblici e, in generale, migliorare la qualità della ricerca, come accade, per fare un solo esempio, nei Paesi Bassi (VSNU KNAW NWO 2020; Morriello 2019).

⁴ Il RAE nel 2014 è divenuto Research Excellence Framework (REF) con l'esercizio di valutazione che copriva gli anni 2008–2013, introducendo alcuni cambiamenti, tra i quali l'enfasi sulla valutazione dell'impatto della ricerca.

La VTR, ispirata al RAE britannico, copriva gli anni 2001-2003 e il rapporto finale fu pubblicato a febbraio 2007. Durante tale esercizio erano stati sottoposti a valutazione 17.329 prodotti della ricerca, distribuiti in 20 aree scientifico-disciplinari (di cui 6 interdisciplinari), presentati da 102 strutture (77 atenei e 25 enti pubblici di ricerca). Per ciascuna delle aree era stato costituito un panel di valutatori nominati dal CIVR, in numero da 5 a 17 a seconda delle aree, per un totale complessivo di 183 pari e 6661 revisori. I panel valutavano i prodotti conferiti esclusivamente tramite la peer review e attribuivano un giudizio tra “eccellente”, “buono”, “accettabile”, “limitato”. Il report finale classificava i prodotti in rapporto all’area disciplinare e li normalizzava in relazione alla dimensione della struttura valutata, definita sulla base del numero di prodotti presentati⁵. I risultati sono stati usati per la prima volta dal Ministero dell’Istruzione, dell’Università e della Ricerca nel 2009 per allocare il 7% del Fondo di Finanziamento ordinario (Abramo, D’Angelo e Caprasecca, 2009; Franceschet e Costantini, 2011).

I costi sostenuti per l’esercizio della VTR⁶ contribuirono ad alimentare anche nel nostro paese il dibattito sull’opportunità di usare la peer review rispetto ai metodi allora considerati meno onerosi e ritenuti meno soggettivi, come la bibliometria. Un dibattito che si collocava in un ampio contesto internazionale di studio e ricerca, molto intenso in quegli anni, e vivace soprattutto in Gran Bretagna, che mirava a rilevare le correlazioni esistenti tra le misurazioni bibliometriche e la peer review. Nel 2011 La House of Commons britannica aveva commissionato uno studio sullo stato della peer review dal titolo *Science and Technology Committee, Eighth Report: Peer Review in Scientific Publications*, nel quale si giungeva alla conclusione che sebbene la peer review sia cruciale per la reputazione e l’affidabilità della ricerca scientifica, molti scienziati ritengono che questo sistema soffochi il progresso scientifico, risulti spesso parziale e vi sia poca evidenza della sua efficacia (Baldwin 2020). Diversi studi in quegli anni attestavano, inoltre, un certo grado di correlazione tra i risultati della peer review e gli indicatori bibliometrici, sebbene le eccezioni rilevate fossero numerose.⁷ Nonostante il principio di base che la correlazione non implica necessariamente un rapporto di causa-effetto, queste analisi ebbero una certa influenza. In generale, la correlazione tra peer review e citazioni se presa a livello di struttura di ricerca e non di indicatori personali (come l’h-index) risultava buona ma in misura variabile a seconda del settore scientifico, mentre la correlazione con le metriche della rivista risultava debole. Le voci a favore della valutazione bibliometrica, in opposizione alla peer review, si fondavano su elementi quali la rilevazione che nella VTR gli atenei non avessero scelto i prodotti migliori, soprattutto in comparazione con i dati bibliometrici, e sostenevano l’inefficacia della peer review per misurare la produttività. Molte voci del dibattito italiano auspicavano, inoltre, la riduzione dei costi e dei tempi dell’esercizio di

⁵ I prodotti conferibili per la valutazione erano nella misura di uno ogni quattro ricercatori di ruolo per ogni ateneo e non superiore al 50% degli FTE dell’istituzione nel periodo considerato.

⁶ Il costo finale dell’esercizio di valutazione VTR è stato stimato in 3,5 milioni di euro.

⁷ Uno studio italiano (Franceschet e Costantini, 2011), per esempio, stabiliva un indice di correlazione intorno allo 0,6% per sei discipline delle otto analizzate, con un massimo di 0,876 per le scienze agricole e veterinarie, ma una correlazione non forte per altri settori come l’ingegneria industriale e la fisica. Franceschet e Costantini individuavano delle debolezze nella correlazione riferita all’impact factor confrontato con i risultati della peer review della VTR in alcuni settori come mathematics and computer sciences (MCS), earth sciences (EAS), physics (PHY), medical sciences (MED), civil engineering and architecture (CEA), industrial and information engineering (IIE), e un disallineamento nella correlazione tra il posizionamento nei quartili e la valutazione VTR (soprattutto nel settore biology).

valutazione che la bibliometria avrebbe consentito rispetto alla peer review, oltre a sostenere diffusamente la convinzione della maggiore oggettività della bibliometria.

Il dibattito acceso anche in Italia sui vantaggi e svantaggi dei due metodi portò all'introduzione della bibliometria nell'esercizio di valutazione successivo alla VTR. Il secondo esercizio di valutazione avrebbe dovuto coprire gli anni 2004-2008 ed era stato ancora assegnato al CIVR ma nel frattempo la struttura per la valutazione della ricerca veniva ridefinita e con il DPR 76 del 1 febbraio 2010 veniva istituita, e se ne definiva il funzionamento, l'Agenzia Nazionale per la Valutazione del Sistema Universitario e della Ricerca (ANVUR). L'ANVUR, insediata il 2 maggio 2011, doveva portare a compimento le attività intraprese e programmate dal CIVR, che cessava di esistere e veniva sostituito dalla nuova agenzia. Il DM del 15 luglio 2011 affidava all'ANVUR lo svolgimento dell'esercizio di Valutazione della Qualità della Ricerca per gli anni 2004-2010 (VQR1). Il bando per la prima VQR è stato pubblicato il 7 novembre 2011. Il rapporto finale è stato diffuso nel 2013. In concomitanza con l'avvio della VQR si cominciava a normare, con la prima di una serie di leggi, l'Abilitazione Scientifica Nazionale (ASN), istituita dall'articolo 16 della Legge n. 240/2010, come modificato dalla Legge n. 114/2014, che attesta la qualificazione scientifica e costituisce un requisito necessario per l'accesso alla prima e alla seconda fascia della docenza universitaria. Il Decreto del Ministero dell'istruzione, dell'università e della ricerca n. 76 del 2012 stabiliva l'introduzione dei criteri bibliometrici per i settori in cui gli indicatori sono disponibili. L'ASN ritornava per certi versi a un istituto che era esistito fino agli anni Settanta, la libera docenza. La libera docenza era un titolo abilitativo, il cui conseguimento autorizzava il titolare (libero docente) all'insegnamento in università e istituti di istruzione superiore (Rossi 2016),⁸ e fu abolita con la legge 924 del 1970, proposta da Tristano Codignola.⁹ Il Decreto del MIUR sull'ASN aveva introdotto la valutazione tramite gli indici bibliometrici e attribuito quindi all'uso di tali strumenti forza di legge. I settori scientifici venivano suddivisi in bibliometrici, quelli facenti capo alle aree da 1 a 9, corrispondenti alle scienze dure, della vita e ingegneristiche, e non bibliometrici, per le aree 10-12 e 14, delle scienze umanistiche e sociali, con l'esclusione delle scienze statistiche ed economiche dell'area 13 valutate con strumenti bibliometrici ma diversi da quelli delle aree 1-9.¹⁰

La bibliometria entrava dunque nella valutazione tramite legge e sarebbe poi adottata per i settori bibliometrici nei successivi esercizi della VQR.¹¹ A tale scelta l'agenzia italiana di valu-

⁸ La libera docenza venne stabilita nella Legge Casati del 1859, promulgata nel Regno di Sardegna e mantenuta in vigore con l'Unità d'Italia. Il Regio Decreto 31 agosto 1933, n. 1592 che regolava la figura del libero docente, lo definiva come un laureato e studioso che superava un esame di abilitazione per titoli. Con tale esame si conseguiva l'abilitazione all'insegnamento per una determinata disciplina, pur in mancanza di una cattedra. Era anche possibile da parte del Ministro per l'educazione nazionale abilitare alla libera docenza «persone che siano venute in alta fama di singolare perizia nella materia che debbono insegnare» (art. 122). La legge distingueva le figure del professore ordinario e professore incaricato.

⁹ Codignola, esponente del Partito Socialista Italiano (dal quale fu espulso nel 1981), fu deputato dal 1958 al 1968 e senatore dal 1968 al 1972. A lui si devono diverse leggi nel campo dell'istruzione e della ricerca come nel 1962 la legge sulla scuola media unica e nel 1969 la legge che liberalizzava l'accesso all'università per tutti i diplomati. Nel 1980 fondò insieme a Giovanni Carocci la casa editrice La Nuova Italia Scientifica che nel 1998 assumerà il nome di Carocci Editore.

¹⁰ Per l'area 13 il GEV aveva predisposto un elenco di riviste e l'articolo veniva valutato in base alla collocazione della rivista nell'elenco, anticipando la classificazione delle riviste in scientifiche e di classe A che verrà successivamente adottata per tutte le aree non bibliometriche.

¹¹ Per i dati sulle VQR si rimanda al sito ANVUR dove sono pubblicati i bandi e la documentazione <https://www.anvur.it/attivita/vqr/>.

tazione ha sempre voluto dare fondamento scientifico, continuando ad alimentare il confronto tra i due metodi di valutazione. Il rapporto dell'ANVUR sulla VQR 2004-2010 (ANVUR 2013) includeva i risultati di un'analisi comparativa tra l'uso degli indicatori bibliometrici e la peer review, compiuta a titolo di verifica sul dieci per cento dei prodotti sottoposti alla valutazione ministeriale, dal quale si traeva la conclusione di una sostanziale corrispondenza tra i due tipi di valutazione. Tale corrispondenza veniva però messa in dubbio da alcuni articoli che, riproducendo l'analisi compiuta da ANVUR,¹² riscontravano invece una discreta corrispondenza tra i risultati della valutazione bibliometrica e della peer review solo in alcuni settori, come Economia e statistica (Baccini e De Nicolao, 2016). In questo settore peraltro il GEV della VQR esprimeva parere positivo all'uso della bibliometria, pur rilevando dalla comparazione con la peer review la prevalenza di un giudizio più favorevole quando scaturito dall'applicazione degli indicatori citazionali rispetto alla revisione tra pari, e suggerendo l'opportunità di svolgere ulteriori analisi comparative (Bertocchi et al. 2015). Anche il secondo esercizio VQR 2011-2014 ha lasciato uno strascico di considerazioni e commenti su diverse questioni come l'uso non appropriato degli indicatori, in particolare dell'impact factor, il non efficace accostamento di metodi di natura diversa come la bibliometria e la revisione tra pari, la scarsa rappresentatività di poche pubblicazioni per valutare l'intera attività scientifica di un docente o ricercatore, l'impossibilità di identificare gli atenei di eccellenza sulla base di un così basso numero di pubblicazioni, la variabile dimensionale degli atenei come bias rispetto ai risultati della valutazione¹³ (Franceschini e Maisano 2017; Benedetto et al. 2017; Abramo 2017; Demetrescu, Ribichini e Schaefer 2020). Il *Rapporto biennale sullo stato del sistema universitario e della ricerca* redatto dall'ANVUR nel 2018 ribadisce l'esistenza di una correlazione positiva tra la valutazione bibliometrica e la peer review come rilevata in alcuni studi, nonostante nello stesso documento vengano citati altri studi che dimostrano l'impossibilità di comparare i due metodi (ANVUR 2018). La VQR 3 ha riguardato gli anni 2015-2019 e la fase di conferimento dei prodotti e si è conclusa ad aprile 2021, dopo essere stata rimandata a causa della pandemia di Covid 19. L'ultimo esercizio di valutazione ha parzialmente accolto i suggerimenti esposti nel rapporto che l'ANVUR stessa aveva commissionato a quattro esperti di valutazione provenienti da Francia, Gran Bretagna, Paesi Bassi e Spagna, per un'analisi dei pro e dei contro delle precedenti VQR (Galderisi et al. 2019). Tra le novità dell'ultima VQR, infatti, vi è stato il rafforzamento della la peer review come strumento base per la valutazione, con la possibilità di supportare il processo di revisione tra pari con i dati bibliometrici, laddove disponibili. I singoli GEV potevano stabilire le proporzioni tra i due metodi. La nuova modalità di valutazione è stata quindi la peer review informata, già prevista nella precedente VQR ma solo per specifici casi.

¹² Con i dati desunti dai risultati finali poiché il dataset delle VQR non era pubblico, nemmeno in maniera anonima, come specificano gli autori dell'analisi.

¹³ Il bias dimensionale è stato confermato anche in uno studio compiuto sui risultati della VQR 2011-2014 alla Sapienza Università di Roma, simulando una partizione dell'ateneo in sezioni più piccole in base a diversi criteri, secondo il quale gli atenei più grossi, come quello romano, sono penalizzati dal metodo utilizzato dalla VQR (Demetrescu, Ribichini e Schaefer 2020).

Peer review e bibliometria

La peer review e la bibliometria sono due modalità di valutazione dalle implicazioni molto diverse. L'una è un surrogato, una misura che, come teorizzava Leo Meltzer, presuppone una correlazione positiva tra la quantità di citazioni e la qualità della produzione scientifica. Un principio spesso smentito dagli studi successivi. Rispetto ai tempi in cui scriveva Meltzer, peraltro, il mondo della ricerca scientifica è profondamente cambiato. Per esempio, l'uso odierno della citazione difficilmente si può definire meno soggettivo della peer review, poiché anche la citazione è legata a variabili come le specifiche esigenze, conoscenze, ambiti di investigazione del ricercatore, i rapporti di collaborazione in atto, le idiosincrasie personali. Come per tutti i dati, anche per i dati bibliometrici occorrerebbe l'analisi del contesto per poterne trarre informazione significativa. Inoltre, la citazione è diventata funzionale alla valutazione e, come avvisa la nota legge di Goodhart, quando una misura diventa un obiettivo cessa di essere una buona misura. Sembra essersi realizzato il timore dei cambiamenti etici che alcuni studiosi paventavano fin dagli albori della bibliometria. Alla metà degli anni '60, Norman Kaplan, ad esempio, ravvisava il rischio che gli indici citazionali potessero causare dei cambiamenti rilevanti nelle pratiche e nell'uso delle citazioni e generare una nuova etica che avrebbe influenzato non solo il modo di pubblicare ma anche i comportamenti citazionali (Kaplan 1965).

Sebbene la peer review presenti indubbiamente un certo grado di soggettività, questo tipo di valutazione si basa sul giudizio complessivo del revisore in grado di differenziare la revisione a seconda delle specificità della disciplina, della tipologia di prodotto, del progetto di ricerca complessivo (Guerrini 2022), e ingloba una serie di parametri testuali e paratestuali che restituiscono un contesto, e che la bibliometria non è in grado di cogliere¹⁴. Per esempio, corredare l'articolo con immagini, grafici o dataset può essere considerato un elemento qualificante ai fini del giudizio finale poiché consente di valutare meglio il metodo e il percorso per raggiungere i risultati, e questo non è rilevabile dagli indici citazionali, così come un articolo composto per la maggior parte da riferimenti bibliografici, invece che dall'esposizione, salterebbe all'occhio del revisore.

La peer review, inoltre, riesce a cogliere la differenza concettuale e semiotica che intercorre tra la citazione e il riferimento bibliografico, sulla quale si sono soffermati diversi autori (Price 1967; Cronin 2005; Wouters 1998 e 2016). Il riferimento bibliografico è un segno che attribuisce un riconoscimento a un altro testo, puntando all'altro testo ma rimanendo contestualizzato nel testo citante ed è dunque un attributo del testo citante. Peraltro, in questo senso, vi è una differenza anche tra i riferimenti bibliografici posti come note a piè di pagina disseminate all'interno di un articolo e i riferimenti bibliografici posti cumulativamente in chiusura dell'articolo. Nel primo caso, inoltre, il segno può puntare sia a una specifica frase sia a una porzione ampia dell'intero testo, e può farlo tramite due tipi di referente costituiti dalla bibliografia finale o dalla nota a piè di pagina del testo citante, oppure dalla porzione di testo citato, laddove presente, assumendo dimensioni significanti diverse. Di converso, la citazione è un attributo del testo citato in quanto è il riconoscimento che un testo riceve da un altro testo. Le funzioni sono differenti e la bibliometria le appiattisce, decontestualizzando e ricontestualizzando in maniera artificiosa le citazioni

¹⁴ Perlomeno nel modo in cui viene utilizzata per la valutazione. L'applicazione della network analysis ai dati bibliometrici consente analisi più profonde e contestualizzate.

secondo logiche diverse dalle originali. Nelle banche dati citazionali il contesto di una specifica citazione non è più legato al testo citante o al testo citato ma, per esempio, all'insieme delle citazioni di una rivista in un determinato arco temporale (per l'impact factor) oppure di un settore scientifico (per gli indicatori pesati sul settore disciplinare come il Field Weighted Citation Index, per esempio). Tali nuove contestualizzazioni sono però fittizie, determinate a posteriori da specifici interventi (dei produttori degli indici citazionali) di aggregazione e distribuzione, e non sono legate ai ritmi, alle reti e alle funzioni epistemiche naturali della citazione all'interno del sistema della ricerca scientifica. Si viene dunque a creare un'astrazione, una rappresentazione della realtà che non corrisponde affatto alla realtà, ma determina un nuovo significato, diverso dalla funzione originale e non in grado di restituire il contesto originale della citazione all'interno del testo. Questo contesto creato artificialmente trasforma la funzione epistemica originaria della citazione, intesa come legame concettuale tra due pubblicazioni, necessario per collegare un articolo alla conoscenza che l'ha generato in senso spaziale e temporale. Come spiega Michael Polanyi, si tratta di un legame fondamentale per l'autoregolamentazione del sistema di coordinamento spontaneo della scienza, basato appunto su relazioni diacroniche e sincroniche che consentono a uno studioso di entrare in relazione con chi sta svolgendo ricerche simili e con chi le ha svolte in passato (Polanyi 1962). Il venir meno di tali fondamenta sgretola uno dei pilastri sui cui si fonda il sistema della scienza.

Le trasformazioni in corso della funzione epistemica della citazione hanno ripercussioni non solo sui settori bibliometrici ma anche sui settori non bibliometrici delle scienze umanistiche e sociali. Da un lato, il venir meno della funzione epistemica della citazione minaccia le capacità di autoregolamentazione della scienza, e questo riguarda tutti i settori scientifici, bibliometrici e non; dall'altro, il rischio che si corre con lo sbilanciamento verso la valutazione quantitativa è che si accentuino le differenziazioni tra le scienze dure e le scienze umanistiche, o meglio che si stigmatizzino le differenze che sono in parte connaturate ai due ambiti, in parte indotte nei decenni dai processi sociali e politici. Come scrive Paola Castellucci «il fatto stesso di pre-definire i metodi di valutazione ripartendoli in bibliometrici/non bibliometrici, ha insinuato che il metodo qualitativo sia definibile solo in termini negativi rispetto a quello quantitativo» e aggiunge «come a dire che per l'area scientifico-tecnica esistono pesi e misure, mentre nell'area umanistica si va avanti in modo impressionistico, giudicando “a occhio”» (Castellucci 2020). Le implicazioni di tale atteggiamento, e di una simile percezione che porta, a volte, in ambiti bibliometrici a definire la questione dei non bibliometrici come “problematica”, non sono limitate alle dinamiche interne delle istituzioni accademiche e della ricerca¹⁵, ma hanno ripercussioni sulla società tutta. Difatti, tale distinzione intensifica quella frattura tra cultura scientifica e cultura umanistica che caratterizza la nostra società e che andrebbe invece ricomposta per affrontare la complessità crescente, rafforzando il dialogo e lo scambio. La pressione generata dal sistema viene percepita dai ricercatori di ambito HSS (Humanities and Social Sciences) anche come conseguenza della tendenza ad applicare i metodi bibliometrici ai settori HSS. Questo genera un altro effetto, ancora non razionalizzato e studiato a fondo, ovvero la propensione dei ricercatori nei settori HSS a modificare le proprie abitudini di ricerca.

¹⁵ Nelle quali si registra peraltro una riduzione dei dipartimenti di area umanistica (Greenfield 2022; Bulaitis 2021; Myklebust 2021).

Data analysis sulle pubblicazioni del SSD M-STO/08 (Archivistica, bibliografia e biblioteconomia) Per raccogliere informazioni utili a supporto dell'ipotesi formulata riguardo alla tendenza al cambiamento nelle pratiche della ricerca, ho condotto un'analisi sui dati delle pubblicazioni presenti in IRIS negli atenei italiani per i diversi settori scientifico-disciplinari, bibliometrici e non bibliometrici.¹⁶ I dati analizzati coprono l'arco temporale 2004-2018, il cui inizio quindi si pone volutamente in corrispondenza con l'inizio di copertura della prima VQR, e sono tratti dalle istanze IRIS degli atenei italiani. IRIS è il CRIS (Current Research Information System) prodotto dal Cineca e in uso nella maggior parte degli atenei italiani, ma con vari gradi di personalizzazione, il che rappresenta il primo limite dei dati che occorre dichiarare. Un altro limite del dataset è la possibilità che non tutti i dati sulle pubblicazioni siano stati inseriti dagli autori nelle istanze IRIS degli atenei, e dunque che i dati non rappresentino l'intera produzione scientifica. Vi è un'altra criticità generale rappresentata dal fatto che non in tutti gli atenei è attivato il processo di validazione dei metadati inseriti in IRIS da parte degli uffici amministrativi o bibliotecari. A volte la compilazione dei campi di IRIS e l'inserimento dei metadati sono responsabilità degli autori delle pubblicazioni e dunque l'uniformità dei dati non è garantita. Potrebbero quindi esserci casi di schede duplicate per la stessa pubblicazione, dati bibliografici non corretti anche rispetto alla data, pubblicazioni catalogate nelle tipologie di prodotto non appropriate. L'inserimento in IRIS presenta alcune ambiguità che ulteriormente inficiano la qualità dei dati, per esempio la possibilità che il campo "manuale didattico" non sia attivo o non sia stato compilato e dunque ci potrebbero essere pubblicazioni presenti nel conteggio che riguardano la didattica e non la ricerca.

Nonostante questi limiti, i dati di IRIS sono importanti perché consentono delle elaborazioni anche per i settori non bibliometrici, poco rappresentati nella banche dati più uniformemente strutturate come Scopus e Web of Science, e principale oggetto di attenzione in questa sede. In linea generale, i risultati dell'analisi dei dati complessivi mostrano alcune tendenze generalizzate per i settori non bibliometrici, con poche eccezioni: un incremento progressivo della tipologia Articolo in rivista in tutti i settori disciplinari (questo vale anche per i settori bibliometrici ma con implicazioni diverse). Sono in diminuzione, invece, i Contributi in atti di convegno e le Monografie, mentre i Contributi in volume sono la tipologia che presenta l'andamento più variabile. I settori non bibliometrici mostrano alcune tendenze comuni, di cui l'SSD M-STO/08 è rappresentativo. Per tale ragione lo specifico interesse del presente contributo per il settore M-STO/08, rispetto al quale proverò a trarre alcune considerazioni generali, offre un'occasione di riflessione per l'intero ambito HSS. L'SSD M-STO/08 fa parte del settore concorsuale 11/A4 Scienze del libro e del documento e scienze storico-religiose. Come per la maggior parte degli SSD, bibliometrici e non bibliometrici, M-STO/08 mostra una linea di crescita del numero di articoli negli anni dal 2004 al 2018 (figura 2).

¹⁶ Il lavoro rientra in un progetto di ricerca condotto nell'ambito del Dottorato in Scienze librarie e documentarie presso la Sapienza Università di Roma che ha riguardato un campione esteso di SSD, bibliometrici e non bibliometrici, non limitatamente a M-STO/08.

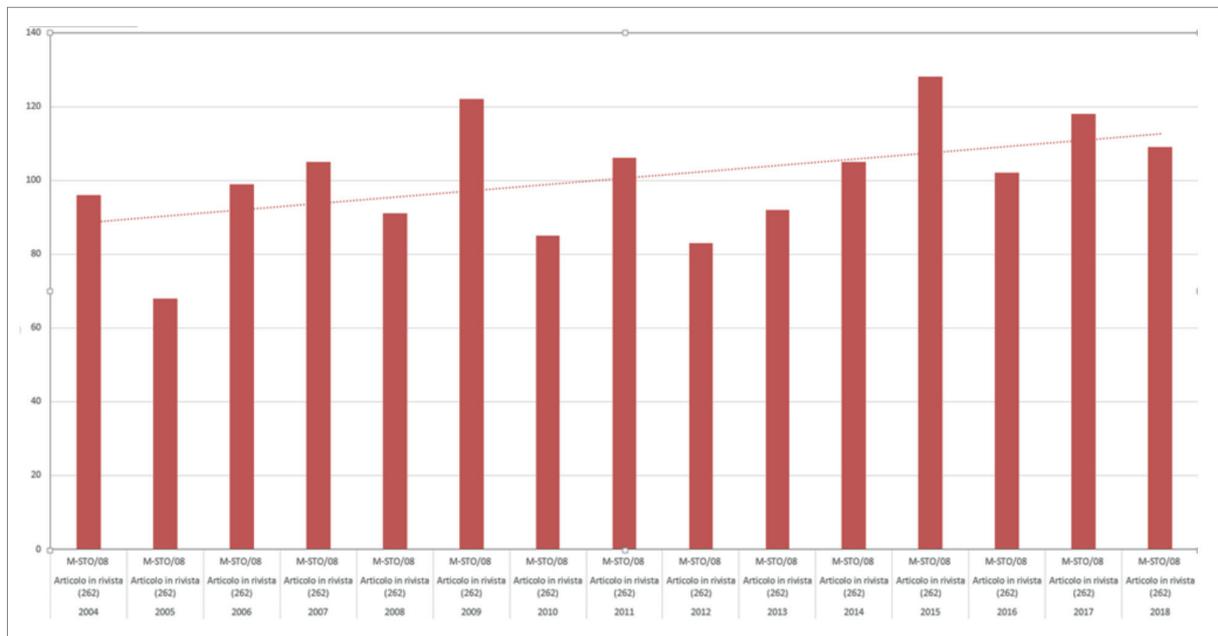


Figura 2. Dati per SSD M-STO/08 e tipologia IRIS Articolo in rivista

La crescita degli articoli che caratterizza M-STO/08 e molti altri SSD, suggerisce che per i settori non bibliometrici vi sia un cambiamento in atto nelle abitudini dei ricercatori. Difatti, mentre nei settori STM l'articolo è la forma di pubblicazione prevalente e l'aumento di tale tipologia non è inatteso, per i settori HSS tale dato suggerisce che i ricercatori si stanno spostando dalla pubblicazione monografica tradizionale e preferiscono suddividere i contenuti che in precedenza andavano a formare una monografia in tanti articoli oppure in contributi in volumi miscelanei. È inevitabile supporre che ciò sia un'esigenza determinata dalla cultura del *publish or perish* e, nello specifico, dalla necessità di superare le soglie quantitative per le varie procedure di valutazione, come l'ASN. In altri termini, sembra concretizzarsi il fenomeno del *salami slicing*, per cui i contenuti che normalmente andrebbero a formare una monografia dopo un lungo percorso di ricerca, anche di diversi anni, vengono frazionati in tanti capitoli in volumi miscelanei o articoli in riviste. In diversi settori non bibliometrici, infatti, la tipologia Contributo in volume è in aumento, oppure è tendenzialmente stabile come nell'SSD M-STO/08 (figura 3).

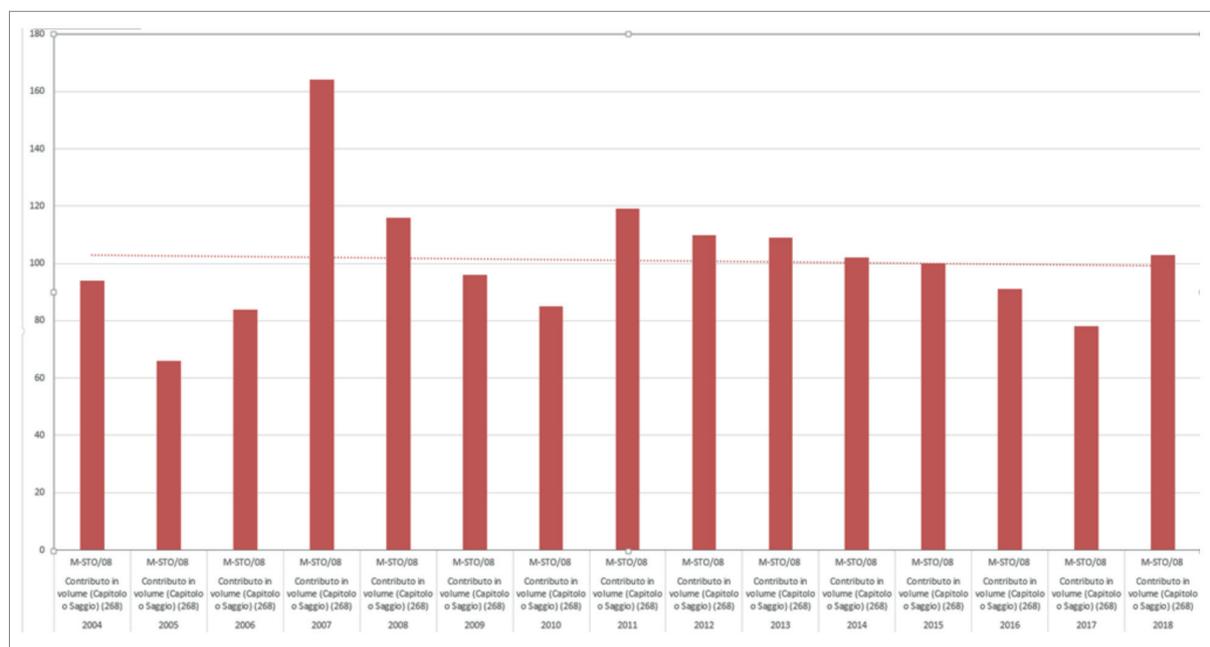


Figura 3. Dati per SSD M-STO/08 e tipologia IRIS Contributo in volume

Le monografie risultano generalmente in diminuzione, ma anche in questo caso se la tendenza è normale per i settori bibliometrici, lo è meno per i settori delle scienze umanistiche e sociali dove la monografia rappresenta tipicamente il principale output della ricerca e dunque anche il riferimento prevalente nell'attività di ricerca. Blaise Cronin sintetizzava le abitudini di pubblicazione dei ricercatori nei diversi ambiti disciplinari, rimarcando come le differenze tra discipline non riguardino solo la produzione di conoscenza ma anche il modo in cui si usa la conoscenza pubblicata e la si cita (Cronin 2005), sostenendo che, ad esempio, per gli storici la monografia interpretativa rimaneva la forma dominante di pubblicazione, nelle scienze economiche era prevalente l'articolo in rivista referato, in fisica il preprint elettronico, nelle scienze informatiche lo era il contributo in atti di convegno. Infatti, come ulteriore conferma della centralità della monografia, un'analisi esposta da Yves Gingras rileva che tre quarti dei riferimenti bibliografici contenuti negli articoli di ambito umanistico fanno riferimento a libri e non ad articoli, mentre in ambito economico si registra già da tempo un calo dei riferimenti ai libri, passati dal 55% al 30%, con una tendenza che li avvicina alle scienze naturali, nelle quali oltre l'80% delle citazioni fa riferimento ad articoli. Le scienze sociali si collocano a metà e fanno riferimento parimenti a libri e articoli (Gingras 2016, 21). Le trasformazioni in atto configurano, dunque, un vero e proprio cambiamento di paradigma nella ricerca umanistica e sociale.

Sembrerebbero confermati i timori evidenziati dal report del Progetto LI.B.RO, lo studio commissionato dall'ANVUR e coordinato da Giovanni Solimine, nell'ambito della valutazione delle aree non bibliometriche che ravvisava il rischio di distorsioni collegato alla valutazione della ricerca nelle HSS, e di fenomeni come il *salami slicing*, soprattutto tra i ricercatori più giovani per i quali la valutazione è strettamente collegata alle possibilità di avanzamento di carriera (Solimine 2017). I settori delle scienze umane e sociali ne sono maggiormente colpiti poiché la monografia è

stata tradizionalmente la forma di pubblicazione tipica di tali settori, come già evidenziato. Il calo delle monografie nei settori STM non sorprende poiché sono gli stessi dati di IRIS analizzati a mostrare come in termini assoluti il numero di monografie sia sempre stato basso in questi settori. La diminuzione delle monografie e l'aumento degli articoli nei settori STM sono dunque fenomeni che non presentano caratteri di novità: aumenta la tipologia di pubblicazione tipica e diminuisce la tipologia meno tipica. Al contrario, un analogo processo per le scienze umanistiche e sociali è degno di attenzione poiché si svolge all'inverso: aumenta la forma di pubblicazione poco tipica e diminuisce la forma di pubblicazione tipica. Un'ulteriore conferma proviene dai dati relativi alle procedure di valutazione presentati da Solimine in un articolo nel quale analizzava e commentava i risultati della VQR 2004-2010, e concludeva riguardo ai prodotti sottoposti alla valutazione che «la 'forma libro' veicola la maggior parte dei prodotti della ricerca: occupa il 55,6 nell'area 10 e il 66,1% nell'area 11, il 61,9% nell'area 12, il 66,9% nell'area 14».49 (Solimine 2016, 243).

I dati generali forniti dall'ANVUR parlano di una quota di monografie sottoposte alla valutazione rispetto al totale dei prodotti presentati pari al 25% nella VQR 2004-2010 e al 20% e nella VQR 2011-2014, aggiungendo la considerazione che «la monografia costituisce l'output di ricerca più prestigioso ed apprezzato in quasi tutte le aree prese in esame», e che «le monografie ricevono voti medi significativamente più alti rispetto alle altre tipologie di prodotti» (ANVUR 2018). La ragione probabile della migliore valutazione delle monografie risiede nel fatto che di norma costituiscono il risultato di un lungo lavoro di ricerca che la forma libro consente di presentare estensivamente, e di cui la valutazione tramite peer review riesce a cogliere i molteplici aspetti, molti dei quali invece non sarebbero rilevabili tramite le misurazioni quantitative e bibliometriche (ammesso che tali dati fossero disponibili sistematicamente per le monografie). In ogni caso, sebbene le monografie siano i prodotti preferiti da sottoporre alla valutazione nelle aree non bibliometriche, il loro numero risulta in diminuzione.

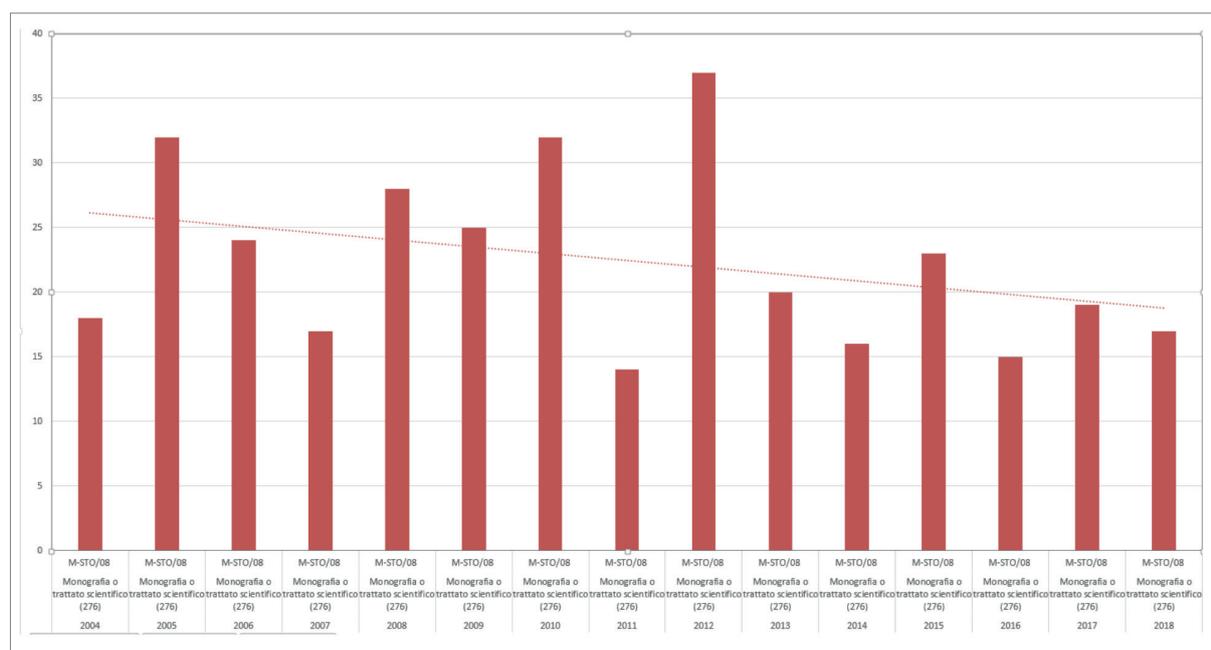


Figura 4. Dati per SSD M-STO/08 e tipologia IRIS Monografia o trattato scientifico

Per quanto riguarda i Contributi in atti di convegno (Proceedings) la tendenza evidente dall'analisi dei dati IRIS è una generale decrescita, sia per i settori bibliometrici sia per i non bibliometrici. Il calo piuttosto importante per il settore M-STO/08 è visibile nella figura 5. In parte ciò potrebbe essere dovuto al diverso trattamento di tale tipologia nei repositories IRIS degli atenei, e da parte degli autori, che danno priorità al contenitore (monografia o periodico) nel quale gli atti dei convegni vengono pubblicati rispetto alla tipologia di contenuto (contributo in convegno). Ma la diminuzione di questa tipologia potrebbe essere dovuta a una effettiva minore partecipazione dei ricercatori ai convegni perché percepiti come meno utili ai fini delle procedure di valutazione e sostituiti da altre forme di trasmissione dei contenuti di una ricerca. Il fenomeno andrebbe approfondito ma se questo fosse il caso, si rischierebbe di perdere, o comunque indebolire, una delle strutture fondamentali della scienza. Le conferenze rappresentano un momento di scambio dei risultati della ricerca, magari dei risultati ancora parziali di ricerche in corso. Per le aree umanistiche svolgono una funzione per certi versi simile a quella che assolvono i preprint nelle aree STM. Sono inoltre occasioni per stabilire contatti e costruire reti di collaborazione, per avviare progetti nazionali e internazionali, e dar luogo a tutte le attività informali che rientrano nella definizione di *invisible college*. L'impatto che le conferenze generano può andare oltre l'immediata pubblicazione degli atti, è sfaccettato e difficile da misurare, e di certo non può essere un indicatore bibliometrico a rappresentarlo. Le aree HSS risentirebbero maggiormente della perdita di valore di tale struttura della scienza.

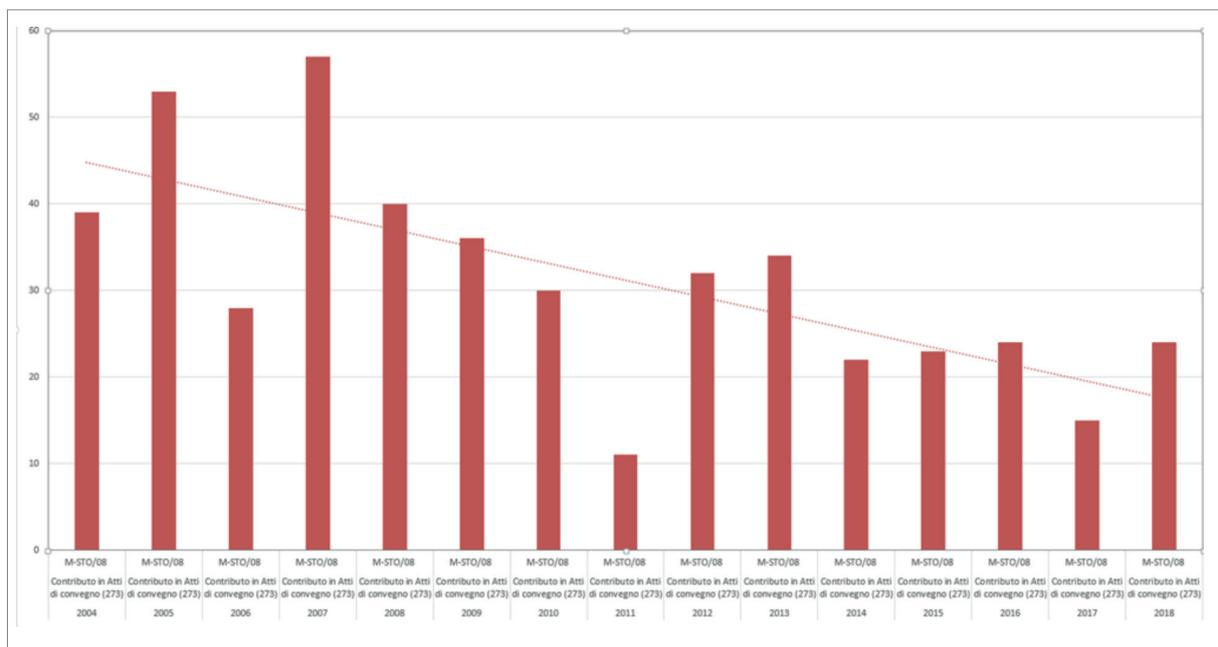


Figura 5. Dati per SSD M-STO/08 e tipologia IRIS Contributo in Atti di convegno

Queste considerazioni e tendenze sono peraltro in gran parte confermate anche dal già citato rapporto ANVUR sullo stato della ricerca del 2018, nonché dai dati estrapolabili dai report della VQR pubblicati dalla stessa agenzia di valutazione sul suo sito, in particolare in relazione alle

tipologie di prodotti conferiti nei due esercizi di valutazione¹⁷. Nella VQR2 la proporzione tra le tipologie Contributo in rivista, Contributo in volume e Monografia scientifica era variabile ma di gran lunga più equilibrata rispetto alla successiva VQR, con una percentuale molto simile tra i tre tipi di prodotti nel 2013, una sostanziale equità tra articoli e contributi in volume pari complessivamente al 50% e con il restante 50% di monografie nel 2012. Si è trattato con probabilità di un periodo di assestamento nel quale non erano ancora rilevabili tendenze precise che emergono invece nel quinquennio successivo. Come si può vedere dai dati riportati nelle tabelle seguenti, i dati della VQR3 mostrano con chiarezza da un lato, la tendenza all'aumento dei contributi in rivista e dei contributi in volume, e dall'altro, la diminuzione delle monografie. L'aumento degli articoli è più marcato, con un raddoppio percentuale tra il 2015 e il 2016 e una crescita rilevante negli anni dopo. Le monografie si sono più che dimezzate dal 2015 al 2019. L'ultima VQR prevedeva, a differenza dell'esercizio precedente, la possibilità di conferire Contributi in atti di convegno, che nell'SSD M-STO/08 sono stati scelti in una percentuale molto bassa, pari al 16% nel 2015 ma poi crollati al 3,13% nel 2016, al 4,88% nel 2017, addirittura a zero nel 2018 e a 1,79 nel 2019. I dati sui prodotti conferiti per le due VQR confermano le tendenze viste nell'analisi dei dati IRIS e lasciano supporre che non si tratti soltanto di una scelta differente nella tipologia di prodotto di conferire per la valutazione ma di un ulteriore effettivo segnale di alcuni cambiamenti in atto, da un lato rispetto al formato preferito di pubblicazione dei risultati della ricerca e dall'altro della percezione di rilevanza dei prodotti da selezionare per un processo di valutazione.

VQR 2011-2014 Prodotti conferiti

SSD	Anno	% Contributo in rivista	% Contributo in volume	% Monografia scientifica
M-STO/08	2011	42,86	28,57	28,57
M-STO/08	2012	25	25	50
M-STO/08	2013	33,33	36,11	30,56
M-STO/08	2014	32,5	47,5	20

VQR 2015-2019 Prodotti conferiti

SSD	Anno	% Contributo in rivist	% Contributo in volume	% Monografia scientifica
M-STO/08	2015	24	16	44
M-STO/08	2016	43,75	31,25	21,88
M-STO/08	2017	58,54	12,2	24,39
M-STO/08	2018	48,21	23,21	28,57
M-STO/08	2019	64,29	16,07	17,86

¹⁷ <https://www.anvur.it/attivita/vqr/>.

Da tale insieme di dati, per i quali sono senz'altro necessari ulteriori approfondimenti, possiamo comunque trarre la conclusione evidente che si stanno producendo delle trasformazioni nelle strutture della scienza e nelle abitudini dei ricercatori, molto probabilmente indotte, perlomeno in parte, dai sistemi di valutazione correnti, non diversamente da quanto accade in altri paesi e da quanto ampiamente discusso nella letteratura scientifica. Le trasformazioni appaiono però di diversa natura tra le aree STM e HSS. Nelle prime i cambiamenti riguardano in misura minore la natura delle strutture e delle pratiche di ricerca. Le aree STM continuano a pubblicare molti articoli, che sono la loro principale forma di output della ricerca, e poche opere di carattere monografico. Nella aree HSS i cambiamenti sono più profondi perché in questi ambiti disciplinari si registra uno spostamento della sede di pubblicazione tradizionale dall'opera di carattere monografico all'articolo su rivista. Lo stesso Rapporto ANVUR conferma che «gli studiosi italiani di scienze umane e sociali, nell'ultimo quindicennio, hanno fatto sempre più ricorso ad articoli su rivista: la quota di questa particolare tipologia di pubblicazione, sulla base dei dati sulla produzione scientifica complessiva, è aumentata costantemente dal 2012 in poi» (ANVUR 2018, 73). Il rapporto conferma poi la tendenza a sottoporre a valutazione di preferenza i prodotti di ambito HSS presenti nella banche dati WoS e Scopus, dichiarando che dalla VQR1 alla VQR 2 sono passati complessivamente dal 10 al 18%, raggiungendo punte del 58% nell'area di Scienze economiche e statistiche (ANVUR 2018).

Le implicazioni di una scelta che privilegia metodi quantitativi, basati su indicatori creati e gestiti da aziende esterne alla comunità scientifica che operano per interesse commerciale, rispetto a uno dei fondamenti del metodo scientifico, e alla base della nascita scienza moderna e delle prime riviste scientifiche fin dal XVII secolo, come la peer review, sono evidenti. Solo tramite il confronto tra pari, che avviene nelle istituzioni scientifiche preposte, come i periodici, i congressi, le altre tipologie di pubblicazioni si può realizzare l'oggettività scientifica necessaria al progresso (Popper 2018). Anche Polanyi riconosceva la centralità dei pari, dei referees, nella validazione dei risultati della ricerca, e del loro giudizio che definisce «opinione scientifica» (Polanyi 2006).

Il futuro della peer review nella valutazione della ricerca

La comunicazione scientifica si basa su un sistema di “arbitraggio” in cui vi sono dei “giudici” che valutano i diversi aspetti delle sue dinamiche (Zuckerman e Merton 1971). Questi “giudici” hanno la responsabilità di aiutare il sistema a mantenere alti gli standard in un settore scientifico e di garantire in questo modo l'avanzamento di tutta la comunità scientifica per quel settore. Anche la peer review usata nella valutazione della ricerca assolve questa funzione di arbitraggio poiché, come abbiamo visto, è divenuta parte del sistema della comunicazione scientifica e i metodi e i risultati della valutazione incidono sui comportamenti dei ricercatori. Occorre dunque chiedersi quale debba essere l'obiettivo della valutazione, se semplicemente la ripartizione di quote di finanziamento, o l'aumento del numero di pubblicazioni, oppure se debba essere un processo volto a migliorare la ricerca scientifica, in modo che possa generare un reale impatto positivo sulla società, che sia inclusiva, sostenibile e solidamente poggiata sulle istituzioni sociali che ne consentono il progresso. Si attribuisce sempre maggiore attenzione alle ricadute che la ricerca produce sulla società, alla Terza e alla Quarta Missione, ma in questa prospettiva dovrebbe rientrare un approccio

olistico, che consideri l'impatto sulla società anche delle forme di valutazione e le distorsioni che possono generare. Per quanto i diversi metodi di valutazione abbiano dei limiti (difficile trovarne di perfetti, in ogni caso), la loro validità andrebbe misurata da una prospettiva ampia e lungimirante, ovvero in termini di effetti negativi più che di effetti positivi, soprattutto sul lungo periodo, e valutando anche il modo in cui si rapportano ai fondamenti della scienza. Il presupposto della scienza è il metodo scientifico, validato per mezzo del giudizio dei pari. Lo evidenziava Popper, secondo il quale “è consigliabile caratterizzare la scienza in base ai suoi metodi piuttosto che in base ai suoi risultati” (Popper 2018, 500).

È ormai chiaro e comprovato da un'ampia letteratura scientifica che la cultura del *publish or perish* indotta dai sistemi di valutazione quantitativa non produce risultati positivi. Difatti, le azioni in atto in molti paesi per invertire la rotta e tornare a una valutazione sostenibile e su basi qualitative sono numerose. La più recente e significativa è l'azione congiunta della European University Association (EUA), di Science Europe¹⁸ e della Commissione Europea che hanno avviato un tavolo di lavoro per riformare la valutazione della ricerca. La bozza di un documento di indirizzo è stata fatta circolare all'inizio del 2022 tra gli stakeholders, promuovendola tramite diversi convegni fruibili in remoto. La bozza è stata aperta ai commenti degli atenei ed altri enti interessati nei primi mesi dell'anno. La versione finale, rivista sulla base dei commenti ricevuti, è stata pubblicata online il 20 luglio 2022 (EUA, Science Europe, Karen Stroobants, e EC 2022). L'*Agreement on Reforming Research Assessment*, che gli atenei e le altre istituzioni sono chiamati a sottoscrivere volontariamente, è molto chiaro nel definire le priorità e individuare il ruolo della peer review:

Our vision is that the assessment of research, researchers and research organisations recognises the diverse outputs, practices and activities that maximise the quality and impact of research. This requires basing assessment primarily on qualitative judgement, for which peer review is central, supported by responsible use of quantitative indicators.

L'*Agreement* non esclude del tutto l'uso degli indicatori quantitativi, se fatto in modo responsabile, a supporto della valutazione, ma ribadisce il ruolo cruciale della peer review più volte nel testo, e sostiene come “peer review is the most robust method known for assessing quality”. Oltre ad essere il metodo più robusto per valutare la qualità, la peer review ha un altro importante vantaggio, ovvero è in mano alla comunità scientifica, ai ricercatori, a differenza di quanto avviene per gli indici bibliometrici che sono invece di proprietà di aziende commerciali private. A ben guardare, sebbene questo sia un aspetto importante poiché le finalità tra aziende private e enti pubblici spesso sono divergenti, ciò su cui occorrerebbe puntare l'attenzione è il modo nel quale le agenzie di valutazione e gli stessi atenei usano tali strumenti quantitativi. Il vero problema è l'adozione di tali indicatori in maniera acritica e distorta rispetto alle finalità per le quali sono prodotti. L'esempio classico è l'uso dell'*impact factor* per valutare i ricercatori, che è esplicitamente sconsigliato dallo stesso *Agreement*. L'*impact factor* è un indice creato per le riviste e non per i singoli articoli o per i ricercatori. Lo stesso creatore, Eugene Garfield, ammoniva di non usarlo per valutare i ricercatori e ne ha riconosciuti i limiti in diverse occasioni (Garfield 1973; 1996). L'*impact factor* totale, in uso

¹⁸ Associazione che riunisce le agenzie di finanziamento della ricerca europee e che per l'Italia include al momento solo l'INFN.

in diversi atenei. è un'ulteriore distorsione applicata non solo a un articolo ma alla più ampia produzione scientifica di un autore. Spesso si attribuisce la responsabilità del ruolo preminente della bibliometria ai produttori degli indici bibliometrici, ma si dimentica che sono gli atenei e le agenzie di valutazione a utilizzarli in maniera non coerente con le finalità per le quali sono stati ideati. *L'Agreement on Reforming Research Assessment* invita a una svolta verso una valutazione complessiva del lavoro del ricercatore, dei diversi tipi di output della ricerca, non solo le pubblicazioni ma anche i brevetti, i dataset, i progetti, e altro, e dell'intero portfolio di attività che includa, oltre alla ricerca, la didattica, il public engagement, gli incarichi istituzionali ricoperti, le buone pratiche sulla scienza aperta e nella condivisione dei risultati della ricerca. Tra queste, auspica anche il riconoscimento dell'attività svolta per la peer review, che è un punto debole in quanto è un'attività onerosa che in genere non ha alcun riconoscimento ai fini della carriera.

Uno dei modi di ridefinire la valutazione della ricerca suggerito dall'Agreement, e già adottato in alcuni paesi, si basa su descrizioni narrative predisposte dal soggetto valutato sulla sua intera attività e mira ad approcciarsi in maniera completamente diversa all'attività di ricerca da valutare. Questo nuovo approccio presuppone una valutazione olistica, sul curriculum vitae completo del soggetto valutato, e viene definito *CV Assessment*. La revisione tra pari va dunque ad estendersi all'intero spettro di attività del ricercatore, assumendo un ruolo ancora più significativo. Per tale ragione si dovrebbero concentrare maggiori energie e risorse per risolvere le attuali criticità della peer review, ampiamente trattate nella letteratura, e gli atenei, i ricercatori, gli editori e le agenzie di valutazione, dovrebbero lavorare in sinergia affinché la peer review, possa essere consolidata e irrobustita, nella fase precedente alla pubblicazione in una rivista, ancor prima che nella valutazione. Alcune iniziative interessanti vanno già in questa direzione, come il Peer Review Portal, il portale australiano che organizza e offre una revisione tra pari indipendente e collaborativa a livello internazionale.¹⁹ Le procedure di valutazione basate sul CV narrativo dovranno includere anche quest'attività onerosa ma fondamentale per la scienza. Naturalmente, si tratta di un obiettivo di non facile realizzazione, perlomeno non allo stesso modo nei diversi paesi europei. Una valutazione di tipo diverso, basata sul CV, richiede un numero più alto di valutatori, competenze estese e tempi più lunghi, il che significa maggiori risorse. In primo luogo, dunque, un investimento specifico su tale attività. L'Agreement non prevede nessun incentivo ma chiede agli stakeholders di destinare risorse specificamente agli obiettivi delineati. Inoltre, solo l'attuazione di una politica di rafforzamento della peer review che passi anche attraverso la sua valutazione ai fini della progressione di carriera potrà consentire una sostanziale trasformazione del processo, come auspicata dall'Agreement.

L'attività della Commissione Europea con EUA e Science Europe, avviata con l'Agreement e che si concretizzerà attraverso l'operato di una Coalizione internazionale,²⁰ avrà auspicabilmente un'influenza importante a livello di diffusione della consapevolezza, poiché è evidente che la posta in gioco è alta: il sistema della ricerca scientifica va sostenuto per far fronte alla complessità sociale crescente, nella quale l'integrità e la qualità della ricerca sono cruciali. È ormai acclarato che i sistemi di valutazione attuali non vadano in questa direzione ed esigano quindi un ripensamento. Tuttavia, l'Agreement non ha nessuna valenza legislativa o regolamentare, ma implica una sem-

¹⁹ Peer Review Portal <https://peerreviewportal.com/>.

²⁰ Coalition for Advancing Research Assessment, <https://www.coara.eu/>.

plice adesione volontaria che, nonostante siano previste delle verifiche periodiche degli adempimenti, non pone nessun vincolo stringente in termini di risultati. Peraltro l'Agreement specifica che non si considera la performance della ricerca a livello nazionale e dunque non si rivolge alle agenzie di valutazione della ricerca, nonostante queste siano comprese tra gli stakeholder e tra i potenziali firmatari. Eppure sono proprio le agenzie di valutazione nazionali a dettare le linee rispetto alla valutazione della ricerca, alle quali poi gli atenei si adeguano e i ricercatori devono sottostare. L'auspicio è che l'iniziativa di EUA, Science Europe ed EC possa comunque rappresentare il primo passo di un percorso di riforma della valutazione che non sarà di certo né breve né facile, ma indispensabile.

Riferimenti bibliografici

- Abramo, Giovanni. 2017. "Correspondence On tit for tat: Franceschini and Maisano versus ANVUR regarding the Italian research assessment exercise VQR 2011–2014". *Journal of Infometrics* 11: 783-787.
- Abramo, Giovanni, Ciriaco Andrea D'Angelo, e Alessandro Caprasecca. 2009. "Allocative efficiency in public research funding: Can bibliometrics help?". *Research Policy* 38: 206-215.
- ANVUR (Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca). 2013. *Rapporto finale sulla VQR 2004-2010*. <https://www.anvur.it/rapporto/>.
- ANVUR (Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca). 2018. *Rapporto biennale sullo stato del sistema universitario e della ricerca*. <https://www.anvur.it/documenti-ufficiali/rapporti-sullo-stato/>.
- Baccini, Alberto, e Giuseppe De Nicolao. 2016. "Do they agree? Bibliometric evaluation versus informed peer review in the Italian research assessment exercise". *Scientometrics* 108: 1651–1671.
- Baldwin, Melinda. 2020. "Peer Review," in *Encyclopedia of the History of Science*. Carnegie Mellon University. <https://lps.library.cmu.edu/ETHOS/article/id/38/>.
- Bence, Valerie e Charles Oppenheim. 2005. "The Evolution of the UK's Research Assessment Exercise: Publications, Performance and Perceptions". *Journal of Educational Administration and History*. 37 (2): 137-155. <https://doi.org/10.1080/00220620500211189>.
- Benedetto, Sergio, Daniele Checchi, Andrea Graziosi, e Marco Malgarini. 2017. "Comments on the paper Critical remarks on the Italian assessment exercise". *Journal of Infometrics*, 11 (2): 622–624.
- Bertocchi, Graziella, Alfonso Gambardella, Tullio Jappelli, Carmela A. Nappi, e Franco Peracchi. 2015. "Bibliometric evaluation vs. informed peer review: Evidence from Italy". *Research Policy* 44 (2): 451-466.
- Bulaitis, Zoe Hope. 2021. "'Minimum expectations' are no way to value the arts, humanities, and social sciences". *LSE Blog*, June 7th. <https://blogs.lse.ac.uk/impactofsocialsciences/2021/06/07/minimum-expectations-are-no-way-to-value-the-arts-humanities-and-social-sciences/>.
- Bush, Vannevar. 2013. *Manifesto per la rinascita di una nazione. Scienza, la frontiera infinita*. Torino: Bollati Boringhieri.
- Castellucci, Paola. 2020. "La visione del giudizio. Una prospettiva romantica," in *Libri, biblioteche e società. Studi per Rosa Marisa Borraccini*, a cura di Alberto Petrucciani, Valentina Sestini, Federico Valacchi. Macerata: EUM.
- Cronin, Blaise. 2005. *The Hand of Science. Academic Writing and Its Rewards*. Lanham: The Scarecrow Press.
- Demetrescu, Camil, Andrea Ribichini, e Marco Schaerf. 2020. "Are Italian research assessment exercises size-biased?". *Scientometrics* 125: 533-549.
- EUA (European University Association), Science Europe, Karen Stroobants, e (EC) European Commission. 2022. *Agreement on Reforming Research Assessment*. https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/reforming-research-assessment-agreement-now-final-2022-07-20_en.

- Faggiolani, Chiara. 2015. *La bibliometria*. Roma: Carocci.
- Franceschet, Massimo, e Antonio Costantini. “The first Italian research assessment exercise: A bibliometric perspective”. *Journal of Infometrics* 5: 275-291.
- Franceschini, Fiorenzo, e Domenico Maisano. 2017. “Critical remarks on the Italian research assessment exercise VQR 2011–2014”. *Journal of Infometrics* 11: 337-357.
- Galderisi, Claudio, Mauro Perretti, Nuria Sebastian Galles, e Thed van Leeuwen. 2019. *Report of the Group of Experts charged by ANVUR to advice on the process ‘Valutazione della Qualità della Ricerca (VQR)’*. An independent assessment on the past VQRs carried out by ANVUR. <http://www.anvur.it/wp-content/uploads/2019/03/High-Experts-Report-on-VQR.pdf>.
- Garfield, Eugene. 1973. “Citation Index depend upon the paper not the journal! Don’t count on ‘citation by association’”, *Current Contents* 22: 5-6. <http://www.garfield.library.upenn.edu/essays/V1p452y1962-73.pdf>.
- Garfield, Eugene. 1996. “How can impact factors be improved?” *British Medical Journal*, 313: 411–413, <https://doi.org/10.1136/bmj.313.7054.411>.
- Geuna, Aldo, e Matteo Piolatto. 2016. “Research assessment in the UK and Italy: Costly and difficult, but probably worth it (at least for a while)”. *Research Policy* 45 (1): 260-271.
- Gingras, Yves. 2016. *Bibliometrics and Research Evaluation: Uses and Abuses*. Cambridge (Ma): MIT Press.
- Greco, Pietro. 2013. “Introduzione,” in V. Bush. *Manifesto per la rinascita di una nazione. Scienza, la frontiera infinita*. Torino: Bollati Boringhieri.
- Greco, Pietro. 2019. *La scienza e l’Europa. Dal secondo dopoguerra a oggi*. Roma: L’Asino D’Oro.
- Greenfield, Nathan M. 2022. “The unkindest cut? – Behind the paring of the humanities”. *University World News*. 13 August 2022. <https://www.universityworldnews.com/post.php?story=20220808152834589>.
- Guerrini, Mauro. 2022. “Sua Maestà il revisore: alcune considerazioni sul processo di peer-review all’interno della LIS”. *AIB Studi*, 61(3): 585–592. <https://doi.org/10.2426/aibstudi-13328>.
- Kaplan, Norman. 1965. “The Norms of Citation Behavior: Prolegomena to the Footnote”. *American Documentation* 16 (3): 179-184.
- Meltzer, Leo. 1956. “Scientific Productivity in Organizational Settings”. *Journal of Social Issues* 12: 32-40.
- Morriello, Rossana. 2019. “La valutazione della ricerca in Olanda. Un’analisi comparativa con il sistema italiano”. *Biblioteche oggi* 37 ottobre: 33-41.
- Myklebust, Jan Petter. 2021. “Copenhagen University set to cut humanities places by 24%”. *University World News*. 20 October. <https://www.universityworldnews.com/post.php?story=20211020154347298>.
- Phillimore, A.J. 1989. “University research performance indicators in practice: The University Grants Committee’s evaluation of British universities, 1985-86”. *Research Policy* 18: 255-171.

- Polanyi, Michael. 1962. "The Republic of Science: Its Political and Economic Theory". *Minerva* 1 (1): 1-32. <https://www.jstor.org/stable/41821153>.
- Polanyi, Michael. 2006. *La società libera. Pensieri liberali*, a cura di Massimo Baldini e Antonello Malavasi. Roma: Armando Editore.
- Popper, Karl. 2018. *La società aperta e i suoi nemici*. Roma: Armando Editore.
- Price, Derek de Solla. 1967. *Sociologia della creatività scientifica*. Milano, Bompiani.
- Pritchard, Alan. 1969. "Statistical bibliography or bibliometrics". *Journal of Documentation* (25): 348-349.
- Rossi, Paolo. 2016. "Stato giuridico, reclutamento ed evoluzione della docenza universitaria (1975-2015)". *RT. A Journal on Research Policy & Evaluation* 1. <https://doi.org/10.13130/2282-5398/6726>.
- Solimine, Giovanni. 2016. "Impatto e valutazione della ricerca scientifica". *Nuovi Annali della Scuola speciale per archivisti e bibliotecari*. 30: 235-251.
- Solimine, Giovanni. 2017. *Progetto LI.B.RO. For a Liable Evaluation of Book's Role in Socio-Economic Sciences and Humanities: an International Comparison*. Report finale. ANVUR, Working Paper 2017/06.
- Villa, Maria Luisa. 2018. *Scienza è democrazia. Come funziona il mondo della ricerca*. Postfazione di Pietro Greco. Milano: Guerini e Associati.
- VSNU (Association of Cooperating Universities in the Netherlands), KNAW (Royal Netherlands Academy of Arts and Sciences), e NWO (Dutch Research Council). 2020. *Strategy Evaluation Protocol (SEP)*. https://www.universiteitenvannederland.nl/files/documenten/Domeinen/Onderzoek/SEP_2021-2027.pdf.
- Weingart, Peter. 2005. "Impact of bibliometrics upon the science system: Inadvertent consequences?". *Scientometrics* 62 (1): 117-131.
- Wouters, Paul. 1998. "The Signs of Science". *Scientometrics* 41 (1-2): 224-241.
- Wouters, Paul. 2016. "Semiotics and Citations," in *Theories of Infometrics and Scholarly Communication. A Festschrift in honor of Blaise Cronin*, edited by Cassidy R. Sugimoto. Berlin: De Gruyter.
- Zuckerman, Harriet, e Robert K. Merton. 1971. "Patterns of Evaluation in Science: Institutionalisation, Structure and Functions of the Referee System". *Minerva* 9 (1): 66-100.