

La VQR di area13: una riflessione di sintesi

Alberto Baccini

Dipartimento di Economia Politica e Statistica
Università degli Studi di Siena

baccini@unisi.it

Sommario

L'esercizio di Valutazione della Qualità della Ricerca 2004-2010 (VQR) condotto dall'Agenzia Nazionale di Valutazione dell'Università e della Ricerca (ANVUR) è stato accolto da critiche che ne hanno sottolineato l'inadeguatezza rispetto agli standard con cui analoghi esercizi sono stati condotti in altri paesi. Nel lavoro si mostra che l'Area 13 è stata uno degli snodi critici più importanti della VQR, dove emergono in modo drammatico problemi metodologici e tecnici tali da rendere i risultati sistematicamente distorti.

1. Introduzione

In molti paesi OECD vengono condotti esercizi nazionali di valutazione della ricerca condotti direttamente dal governo o da altre istituzioni, con lo scopo di verificare lo stato di salute dei sistemi nazionali di ricerca, e permettere l'adozione di politiche informate di finanziamento statale. La pratica fu adottata per la prima volta dal governo britannico di Margaret Thatcher che nel 1986 condusse il primo RAE (Research Assessment Exercise). L'Italia svolse il primo esercizio di valutazione triennale della ricerca (VTR) della ricerca ben 15 anni dopo (A. Baccini, 2013b), ispirandosi proprio a quel modello, e dando luogo ad un'esperienza condotta secondo standard ragionevoli, costosa (A. Baccini et al., 2013), con alcuni difetti che avrebbero dovuto essere superati nelle applicazioni successive (Emanuela Reale, 2008). Solo nel luglio 2011 quell'esperienza ha avuto un seguito con l'inizio della Valutazione della Qualità della Ricerca 2004-2010 (da ora in poi VQR) guidata dalla allora neonata Agenzia Nazionale di Valutazione della Ricerca (ANVUR). La VQR è apparsa fin da subito come profondamente diversa sia rispetto al precedente esercizio di valutazione, che rispetto alle metodologie adottate in ambito internazionale. Queste innovazioni sono state accolte da rilievi critici che ne hanno sottolineato l'inadeguatezza rispetto agli standard prevalenti a livello internazionale (Baccini, 2012a, Banfi, 2012, Coin, 2012). Nel mio intervento intendo riprendere e sottolineare alcune di queste critiche. In particolare intendo mostrare che la valutazione di Area 13, cioè la valutazione delle scienze economiche e statistiche, è stata uno degli snodi critici più importanti della VQR, dove sono emersi in modo esemplare problemi metodologici

e tecnici tali da rendere i risultati sistematicamente distorti. Dopo aver discusso brevemente le più rilevanti criticità, raccogliendole in quattro categorie distinte, concludo con alcune proposte di revisione in grado di eliminare le principali distorsioni.

2. La composizione del GEV13: un problema di *fairness*

Il primo elemento critico della VQR riguarda in generale l'assenza di trasparenza nelle procedure di nomina del personale chiamato a svolgere l'esercizio di valutazione. In (Baccini, 2013a) sono discussi in dettaglio i problemi relativi alla *governance* complessiva dell'esercizio, per cui è possibile soffermarci qui sulla sola questione della composizione dei Gruppi di esperti di valutazione (GEV). I GEV, uno per ogni area di ricerca, sono lo snodo cruciale della VQR perché hanno svolto tutte le attività connesse direttamente alla valutazione della ricerca. In particolare hanno definito i criteri di valutazione bibliometrica, hanno definito le procedure per decidere quali lavori dovessero essere valutati con la bibliometria e quali con la *peer review*, hanno scelto i revisori, li hanno coordinati, hanno sintetizzato i rapporti di revisione, hanno valutato in molti casi direttamente i lavori sottoposti a valutazione. La delicatezza di tali attività avrebbe dovuto suggerire ad ANVUR di prendere molto sul serio la questione della nomina dei membri dei GEV. Se si prende il caso del RAE britannico la composizione dei panel di esperti è sottoposto ad uno scrutinio attento per valutarne l'equità (*fairness*) della composizione in termini di genere e competenze (RAE, 2004). La *fairness* nella composizione di un panel di esperti è funzionale

alla rappresentazione di punti di vista e approcci metodologici diversificati (Baccini and Ricciardi, 2012).

In Italia il Decreto ministeriale (DM MIUR n. 17 del 15/07/2011) che ha governato la VQR era molto laconico, limitandosi a prevedere che fosse il consiglio direttivo di ANVUR a nominare i membri dei GEV ed a scegliere tra loro il presidente. Il consiglio direttivo di ANVUR non ha seguito neanche questa regola ed ha nominato prima i presidenti dei GEV (il file con i nomi dei presidenti è datato 10 ottobre 2011 ed è stato reso pubblico intorno a quella data), e dopo quasi due mesi i membri dei GEV (12 dicembre 2011). I presidenti dei GEV, come dichiarato da uno dei membri del consiglio direttivo di ANVUR, sono stati consultati nella fase di stesura delle regole di funzionamento della VQR (Anonymous, 2011); è pertanto verosimile che abbiano anche avuto voce in capitolo nella scelta degli altri membri dei GEV.

Nel caso del GEV 13 non sono state rispettate regole elementari di fairness per quanto riguarda la composizione di genere, con appena il 16,7% di donne – che hanno poi scelto revisori donne nel solo 20% dei casi (ANVUR, 2013c). E soprattutto il GEV è risultato composto da studiosi strettamente collegati tra loro da rapporti di coautoraggio, come risulta dalla Figura 1. L'assenza di fairness nella composizione del panel ha probabilmente ridotto al minimo le voci di dissenso rispetto ai metodi e alle regole di valutazione adottate dal GEV.

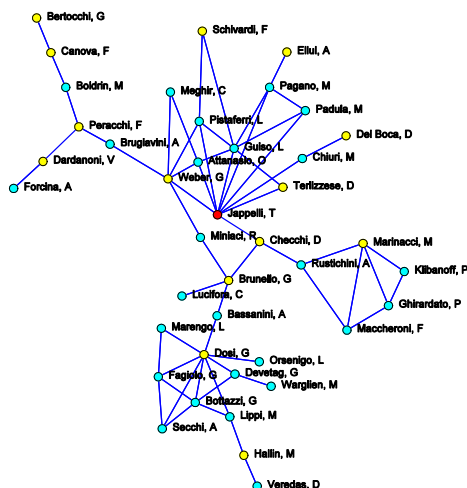


Figura 1. La rete di coautoraggio del GEV di Area 13. Fonte: (Baccini, 2011). Legenda: In rosso il presidente del GEV; in giallo i membri del GEV, in celeste coautori di membri del GEV che non fanno parte del GEV. Una linea che unisce due nomi indica che i due studiosi hanno firmato insieme almeno un articolo.

3. L'insostenibile inconsistenza della valutazione bibliometrica basata sulla classifica delle riviste

Il GEV di Area 13 è l'unico che, sulla base di una presunta specialità delle discipline economiche e statistiche, ha ritenuto di adottare una bibliometria *sui generis* che non trova riscontro in nessun altro esercizio di valutazione condotto a livello internazionale. In particolare la valutazione degli articoli su rivista è stata condotta sulla base di una classificazione delle riviste sviluppata dallo stesso GEV. Anche in questo caso va notata anzitutto l'anomalia procedurale di un gruppo di esperti che si sostituisce alla comunità scientifica nel definire la lista delle riviste e la loro classificazione in fasce di qualità. Di norma questo compito è riservato alle comunità scientifiche che attivano procedure più o meno complesse di consultazione degli studiosi, come avvenuto per esempio in Francia con le liste di riviste di economia sviluppate dal CNRS.

In secondo luogo va notato che l'uso di liste di riviste come surrogato della valutazione bibliometrica diretta degli articoli è una procedura che dà luogo a risultati che sono considerati generalmente privi di validità. Possiamo limitarci a ricordare a questo proposito che la San Francisco Declaration on Research Assessment (DORA, 2012), ad oggi sottoscritta da circa 550 organizzazioni scientifiche di tutto il mondo, stigmatizza l'uso improprio di misure di impatto delle riviste –principalmente l'Impact Factor– come surrogato della valutazione di impatto e qualità degli articoli e dei ricercatori. Una presa di posizione del tutto simile è stata adottata dal board dell'Institute of Electrical and Electronics Engineers (IEEE), la più grande organizzazione internazionale di studiosi di tecnologie, in una dichiarazione sull'uso corretto degli indicatori bibliometrici dove si legge che “*the use of the bibliometric index of a journal in which a researcher publishes (typically the Impact Factor (IF) as a proxy for the quality of his/her specific paper is a common example of a technically incorrect use of bibliometrics*” (Boards of Directors IEEE, 2013). Per una rassegna sintetica e aggiornata sul tema si rimanda a (Brembs et al., 2013; Brembs, 2013).

In Area 13 il GEV ha addirittura messo a punto un'inedita metodologia di classificazione delle riviste, considerata statisticamente inconsistente (Torelli, 2014). È inoltre da notare che il GEV 13 ha fatto affidamento su Google Scholar generalmente considerato allo stato inutilizzabile a fini di valutazio-

ne perché “[it] is not built from structured records, that is from metadata fields. Rather than using the author, affiliation, reference etc. data provided by publishers, Google Scholar parses full-text to obtain its best guess for these items.” (Hicks and Wang, 2009). Sul punto si può vedere anche (Baccini, 2012b; Delgado López-Cózar et al., 2014; Gambardella, 2012; Labbé, 2010).

Last but not least, la classificazione delle riviste non adotta alcuna normalizzazione in relazione alle comunità di riferimento delle riviste. È ben noto che gli indicatori citazionali delle riviste sono sensibili alle modalità di comunicazione e citazione prevalenti nelle diverse comunità scientifiche (Zitt, 2010). Ciò rende particolarmente difficile operare comparazioni tra riviste di settori diversi: per esempio in area medica le riviste di oncologia hanno indicatori più elevati di quelle di epidemiologia che a loro volta hanno indicatori migliori delle riviste interdisciplinari; analogamente in matematica le riviste che pubblicano articoli applicati sono molto più citate delle riviste di matematica pura che a loro volta sono molto più performanti delle riviste di storia della matematica. Per far fronte a questo tipo di problemi, e per permettere la comparazione tra riviste di campi diversi sono stati sviluppati indicatori bibliometrici specifici, tra i quali forse il più noto è il Source Normalized Impact per Paper (SNIP) (Waltman et al., 2013). Anche in questo caso anziché utilizzare la migliore tecnologia bibliometrica disponibile, il GEV 13 ha preferito una soluzione inedita (Maccabelli, 2012), adottando indicatori distorti settorialmente.

4. Bibliometria, *peer review* o entrambe? Un esperimento fallito

Gli esercizi di valutazione possono essere classificati secondo due modelli. Esercizi basati sulla *peer review*, come il RAE/REF britannico. Ed esercizi basati su un modelli bibliometrici puri, del tipo dell'ERA australiano. ANVUR ha ritenuto di perseguire un modello ibrido che usa sia *peer review* che bibliometria per valutare i singoli lavori di ricerca. Mentre le scienze umane, sociali e la giurisprudenza hanno fatto ricorso esclusivo alla valutazione *peer*, nelle scienze dure – le cosiddette aree bibliometriche corrispondenti alle aree CUN 1-9- e in Area 13 alcuni lavori sottoposti a valutazione sono stati valutati da pari, ed altri, la schiacciante maggioranza, con indicatori bibliometrici (ANVUR, 2013b).

Per giustificare questa scelta ANVUR ha altresì condotto un “esperimento”, sottoponendo il 5% circa degli articoli su rivista a una doppia valutazione, sia bibliometrica che attraverso *peer review*, per verificare la concordanza dei risultati tra le due tecniche. I risultati presentati in appendice al rapporto finale indicano chiaramente che i risultati che si ottengono con le due tecniche sono diversi (ANVUR, 2013a). In tutte le aree bibliometriche il grado di concordanza (kappa di Cohen) calcolato da ANVUR è compreso tra 0,17 e 0,35, che significa (Richard Landis and Koch, 1977) un grado di concordanza dei risultati delle due tecniche compreso tra limitato e moderato. Area 13 rappresenta un'eccezione con un indice di concordanza relativamente elevato (0,68).

Questo risultato anomalo di Area13 è frutto di un “esperimento” realizzato con modalità radicalmente difformi rispetto a quanto accaduto nelle altre aree bibliometriche. In estrema sintesi (per una analisi di dettaglio si veda (Baccini, 2014)):

1. Il campionamento degli articoli in Area 13, a differenza di quanto avvenuto nelle altre aree, non è stato casuale, poiché è stato aggiustato tenendo conto delle richieste di accedere alla *peer review* fatte da parte degli autori degli articoli;
2. I revisori di Area 13, a differenza dei revisori delle altre aree, erano nelle condizioni di sapere che stavano partecipando alla sperimentazione;
3. I revisori di Area 13, a differenza dei revisori delle altre aree, erano nelle condizioni di conoscere con estrema precisione il giudizio derivante dall'applicazione delle bibliometria (era sufficiente conoscere la sede di pubblicazione per avere immediatamente il giudizio bibliometrico);
4. In Area 13, a differenza di quanto avvenuto nelle altre aree, il giudizio dei due revisori non è stato sintetizzato sulla base di un algoritmo (media dei punteggi). È stato invece definito da un gruppo di consenso composto da (almeno) due membri del GEV, che ha considerato i due giudizi dei revisori come mere informazioni per definire il giudizio finale;
5. I membri del GEV che hanno costituito i gruppi di consenso sapevano che il loro giudizio finale sarebbe stato utilizzato per la sperimentazione;
6. Di fatto, adottando ipotesi molto conservative, si può calcolare che non meno del 53% dei giudizi derivanti da *peer review* non è stato espresso dai revisori, ma direttamente dai membri del GEV13.

Queste condizioni non solo hanno modificato l'esperimento in modo tale da alterarlo significativamente rispetto a quanto avvenuto nelle altre aree, ma sembra di poter affermare che sono state le procedure a determinare il risultato finale. Di fatto l'elevato indice di concordanza calcolato per Area13 indica che la valutazione basata sulla lista di riviste sviluppata dal GEV 13 si accorda con i giudizi che i membri del GEV 13 hanno dato quando essi hanno deciso le valutazioni nei Gruppi di consenso.

5. I risultati finali dipendono dal "mix valutativo"

Giuseppe De Nicolao (2014) ha mostrato che nelle aree bibliometriche, l'uso di indicatori bibliometrici differenziati per area e settore disciplinare ha reso i risultati dell'esercizio di valutazione non comparabili tra aree e tra settori disciplinari. Il diverso mix di peer review e bibliometria adottato nelle diverse aree, ha giocato un ruolo determinante nell'indurre distorsioni. In particolare, come riconosciuto dalla stessa ANVUR, i risultati delle valutazioni bibliometriche sono stati in media "più generosi" dei risultati raggiunti con la peer review. Per avere un'idea dell'entità della distorsione può essere utile confrontare i risultati VQR con quelli VTR 2001-2004. L'esercizio VTR utilizzava la sola peer review come strumento di valutazione dei lavori di ricerca, ed era ispirato da una logica completamente diversa (CIVR, Comitato di Indirizzo per la Valutazione della Ricerca and MIUR, Ministero dell'Istruzione dell'Università e della Ricerca, 2006; Reale, 2008). Venne infatti chiesto ad ogni struttura valutata di sottoporre a valutazione i migliori prodotti della ricerca. Per ogni struttura era definito il numero massimo di lavori da sottoporre a valutazione, ma non faceva alcuna differenza se i lavori provenivano da molti ricercatori o soltanto da uno. Nell'esercizio VQR è stato invece richiesto di inviare tre prodotti per ogni ricercatore che facesse parte della struttura valutata. La VQR è palesemente rivolta a stimare la "qualità media" della ricerca di una struttura, mentre la VTR era invece rivolta a valutare la presenza di ricerca eccellente nella struttura.

Per capire la differenza può essere utile confrontare le modalità di selezione da parte di un dipartimento dei lavori da sottoporre a valutazione nei due esercizi. Immaginiamo per semplicità un dipartimento tenuto ad inviare a VTR e VQR uno stesso numero di prodotti. Per la VTR il direttore agì in questo modo: ordinò secondo qualche criterio di qualità

tutti i prodotti di ricerca di tutti i docenti del dipartimento, dal migliore al peggiore. Scelse quindi i migliori della lista fino ad arrivare al numero richiesto. Nella VQR invece il direttore ha chiesto a ciascuno ricercatore del suo dipartimento di inviare i suoi tre lavori migliori. Se un ricercatore particolarmente bravo e produttivo nel periodo considerato avesse prodotto un numero di lavori eccellenti superiore a tre, il direttore non avrebbe avuto modo di sottoporre i lavori eccellenti eccedenti i tre alla valutazione, come accade invece per la VTR.

Da questo segue che, proprio a causa del disegno dei due diversi esercizi di valutazione, ci si dovrebbe aspettare, per ogni area scientifica, che la quota di lavori eccellenti rilevata con la VQR sia inferiore a quella rilevata con la VTR. La Tabella 2 confronta la quota di lavori valutati eccellenti nella VTR e nella VQR per area CUN.

Come si può notare nelle aree CUN 10, 11, 12 e 14, cioè nelle aree dove la valutazione è stata condotta con la sola *peer review*, la quota di prodotti eccellenti è drasticamente più bassa nella VQR che nella VTR. Nelle scienze giuridiche (Area12), per esempio, la quota di lavori giudicati eccellenti è scesa dal 20% della VTR all'11% nella VQR. Nella aree bibliometriche è successo invece l'opposto, con risultati eclatanti nell'area dell'ingegneria (Area 9) dove, a causa di ulteriori distorsioni nella procedura (Baccini and De Nicolao, 2012), i prodotti eccellenti sono passati dal 21% della VTR al 51,8% della VQR.

	Percentuale di lavori giudicati eccellenti nella VQR	Percentuale di lavori giudicati eccellenti nella VTR	Differenza tra percentuali VQR-VTR
Area1	44,8%	36,0%	8,8%
Area2	57,2%	52,0%	5,2%
Area3	56,8%	31,0%	25,8%
Area4	30,2%	34,0%	-3,8%
Area5	41,6%	33,0%	8,6%
Area6	38,4%	25,0%	13,4%
Area7	43,0%	10,0%	33,0%
Area8 a	43,2%	22,0%	21,2%
Area8 b	5,2%	22,0%	-12,8%
Area9	52,8%	21,0%	31,8%
Area10	24,2%	51,0%	-26,8%
Area11 a	16,0%	28,0%	-12,0%
Area11 b	35,8%	28,0%	7,8%
Area12	11,2%	20,0%	-8,8%
Area13	19,3%	17,0%	2,3%
Area14	8,9%	20,0%	-11,1%

Tabella 2. Prodotti giudicati eccellenti nella VQR e nella VTR. Un confronto.

Fonte: Elaborazioni su dati (ANVUR, 2013b) e (CIVR, Comitato di Indirizzo per la Valutazione della Ricerca, 2006)

Da queste considerazioni segue una conseguenza non certo secondaria sui risultati finali della VQR. Il voto finale attribuito alla struttura, all'area o ai

settori scientifici valutati è stato sistematicamente più basso tutte le volte che si è fatto ricorso relativamente più forte alla peer review. Di fatto i risultati medi di strutture/aree/settori riflettono semplicemente il diverso mix di bibliometria/peer review che è stato usato nella valutazione dei prodotti di strutture/aree/settori. Le strutture/aree/settori con un maggior numero di prodotti valutati attraverso peer review hanno ottenuto risultati medi sistematicamente più bassi.

In particolare in Area 13 i risultati in termini di voto medio conseguiti dai SSD è inversamente correlato (-0,77) alla quota di prodotti valutati con peer review, come si vede dal grafico 1.

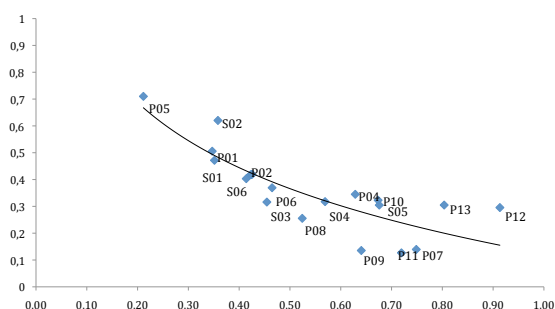


Grafico 1. Voto medio dei settori scientifico disciplinari (SSD) di Area 13 e quota di prodotti giudicati con peer review. Fonte: Elaborazioni su dati (ANVUR, 2013b)

6. Conclusioni

I risultati della VQR sono il frutto di una procedura macchinosa e complessa che ha introdotto tali e tanti elementi distorsivi, da renderli del tutto inutilizzabili a fini conoscitivi e di *policy*. Questo è vero per tutte le aree scientifiche in cui è stata organizzata la VQR, ed in particolare per l'Area 13. Per rimediare agli errori commessi non è possibile intervenire con semplici correttivi sulla struttura VQR, ma è necessario un ripensamento dell'intero modello.

L'adozione del modello britannico REF avrebbe alcuni indubbi vantaggi rispetto alla VQR. Il principale è che l'uso sistematico ed esclusivo della *peer review* insieme ad una corretta definizione di una unica scala di qualità dei prodotti, permetterebbero di salvaguardare la comparabilità dei risultati tra settori e aree diverse.

Nel modello REF è inoltre chiarissima la responsabilità della valutazione che è completamente in capo ai membri del panel. Ogni lavoro viene infatti valutato da un solo revisore membro del panel o designato direttamente e nominativamente da un

membro del panel.

Questo modo di procedere rende delicatissima la questione della composizione dei gruppi di esperti. Per la loro costituzione si dovrebbero seguire procedure pubbliche trasparenti. La composizione del panel dovrebbe permettere la copertura di gran parte dei settori disciplinari di area. La composizione dei GEV dovrebbe altresì rispettare elementari principi di *fairness* per quanto riguarda genere, età e provenienza geografica dei membri del panel. Anche in questo caso sarebbe auspicabile l'adozione del modello di analisi britannico per la valutazione della corretta composizione del panel.

Sarebbe altresì opportuno adottare qualche regola aggiuntiva che impedisca quanto documentato per il GEV di Area13. Si potrebbe per esempio adottare una semplice regola per cui non più del 10% dei membri di un panel possono essere co-autori tra loro o avere un co-autore comune. Si dovrebbe inoltre prevedere che nessun presidente GEV, e nessuno dei membri GEV della VQR 2004-2010 possa far parte dei GEV nel nuovo esercizio di valutazione.

Molte dei problemi illustrati in questo articolo erano stati sollevati prima dell'inizio della scorsa VQR, per cui non c'è da essere ottimisti sul futuro della valutazione della ricerca in Italia.

Riferimenti bibliografici

Anonymous, 2011, "La Ricerca Degli Atenei Italiani All'esame Dell'anvur," *Pisainforma.it*, Pisa, Comune di Pisa,

ANVUR, 2013a, "Rapporto Finale Vqr. Appendice B il Confronto Tra Valutazione Peer e Valutazione Bibliometrica," http://www.anvur.org/rapporto/files/Appendici/VQR2004-2010_AppendiceB.pdf.

_____. 2013b, "Rapporto Finale. Valutazione Della Qualità Della Ricerca 2004-2010", http://www.anvur.org/rapporto/files/VQR2004-2010_RapportoFinale_parteprima.pdf, Roma.

_____. 2013c, "Valutazione della Qualità della Ricerca 2004-2010 (Vqr 2004-2010), Rapporto Finale di Area, Gruppo di Esperti della Valutazione dell'Area 13 (Gev13)", Roma.

Baccini, A., 2013a, "Come e Perché Ridisegnare la Valutazione", Il Mulino, Bologna, 80-87.

_____. 2011. "Gli Esperti Di Valutazione All'italiana," <http://www.roars.it/online/gli-esperti-di-valutazione-allitaliana/>.

_____, 2012a, "La Valutazione Della Ricerca in Ita-

- lia”, *Scuola democratica Learning for democracy*, 5, giugno, 5.
- _____, 2013b, “La Valutazione Della Ricerca in Italia, in ritardo e tecnicamente inadeguata”, *Ricerca*, (5/6), 24-27.
- _____, 2014, “Lo Strano Caso delle Concordanze della Vqr,” <http://www.roars.it/online/lo-strano-caso-delle-concordanze-della-vqr/>. www.roars.it.
- _____, 2012b, “Primo Capitolo ed il suo H-Index”, <http://www.roars.it/online/primo-capitolo-e-il-suo-h-index/>.
- Baccini, A., F. Coin and G. Sirilli, 2013, “Costi e Benefici della Valutazione della Ricerca e della Didattica”, *Paradoxa*, VII (2), 49-61.
- Baccini, A. and G. De Nicolao, 2012, “Vqr: Tutte le Valutazioni sono Uguali, ma alcune sono più Uguali delle Altre,” www.roars.it, <http://www.roars.it/online/vqr-tutte-le-valutazioni-sono-uguali-ma-alcune-sono-piu-uguali-delle-altre/>.
- Baccini, A. and M. Ricciardi, 2012, “Vqr, la Composizione dei Gev ed una Questione di Fairness”, <http://www.roars.it/online/vqr-la-composizione-dei-gev-ed-una-questione-di-fairness/>.
- Banfi, A. 2012. “Qualche Osservazione Su Valutazione Della Ricerca E Scienze Umane.” *Bullettino dell’Istituto di Diritto Romano “Vittorio Scialoja”*, 106(1), 437-62.
- Brembs, B., K. Button and M. Munafò, 2013, “Deep Impact: Unintended Consequences of Journal Rank”, *Frontiers in Human Neuroscience*, 7.
- Brembs, B., 2013, “What Ranking Journals Has in Common with Astrology”, *RT. A Journal on Research Policy and Evaluation*, 1(1).
- CIVR, Comitato di Indirizzo per la Valutazione della Ricerca, 2006, “Relazione Finale Vtr 2001-2003,” Roma, CIVR-MIUR,
- CIVR, Comitato di Indirizzo per la Valutazione della Ricerca and MIUR, Ministero dell’Istruzione dell’Università e della Ricerca, 2006, “Linee Guida Per La Valutazione Della Ricerca,” Roma.
- Coin, F., 2012, “La Valutazione: da Ranking a Profilo di Qualità”, *Scuola democratica*, (5), 160-64.
- De Nicolao, Giuseppe, 2014, “Vqr da Buttare? Persino Anvur Cestina i Voti Usati per l’assegnazione Ffo 2013”, <http://www.roars.it/online/vqr-da-buttare-persino-anvur-cestina-i-voti-usati-per-lassegnazione-ffo-2013/>.
- Delgado López-Cózar, E., N. Robinson-García and D. Torres-Salinas, 2014, “The Google Scholar Experiment: How to Index False Papers and Manipulate Bibliometric Indicators”, *Journal of the Association for Information Science and Technology*, 65(3), 446-454.
- DORA, 2012, “San Francisco Declaration on Research Assessment. Putting Science into the Assessment of Research”, <http://am.ascb.org/dora/>.
- Gambardella, M., 2012, “Google Scholar: Soluzione O Problema?”, <http://www.roars.it/online/google-scholar-soluzione-o-problema/>.
- Hicks, D. and J. Wang, 2009, “Towards a Bibliometric Database for the Social Sciences and Humanities”, Georgia Institute of Technology,
- IEEE, Boards of Directors, 2013. “Appropriate Use of Bibliometric Indicators for the Assessment of Journals, Research Proposals, and Individuals”, http://www.ieee.org/publications_standards/publications/rights/ieee_bibliometric_statement_sept_2013.pdf.
- Labbé, C., 2010, “Ike Antkare One of the Great Stars of the Scientific Firmament”, *ISSI newsletter*, 6(2), 48-52.
- Landis, J. R. and G. G. Koch, 1977, “The Measurement of Observer Agreement for Categorical Data”, *Biometrics*, 33(1), 159-174.
- Maccabelli, T., 2012, “L’inedita Bibliometria di Area 13: Valutazione e Ranking delle Riviste Economiche”, <http://www.roars.it/online/inedita-bibliometria-dellarea-13-valutazione-e-ranking-delle-riviste-economiche/>. www.roars.it.
- RAE, 2004, “Units of Assessment and Recruitment of Panel Members”, London, HEFCE-RAE.
- Reale E. (ed.), 2008, *La Valutazione della Ricerca Pubblica. Una Analisi della Valutazione Triennale della Ricerca*, Milano, Angeli.
- Torelli, N., 2014, “Bibliometria: Fin Dove Arrivare? Alcune Riflessioni su Vqr e Asn per l’Area 13,” Roma, CUN, 6 giugno.
- Waltman, L., N. J. van Eck, T. N. van Leeuwen and M. S. Visser, 2013, “Some Modifications to the Snip Journal Impact Indicator”, *Journal of Informetrics*, 7, 272-285.
- Zitt, M., 2010, “Citing-Side Normalization of Journal Impact: a Robust Variant of the Audience Factor”, *Journal of Informetrics*, 4, 392-406.