Performance-based incentives, research evaluation systems and the trickle-down of bad science

By Alberto Baccini[1]

Governments or university managers, vice-chancellors or rectors need information about effectiveness of research funding. Governments are increasingly using monitoring tools as instruments for governance. The promise of bibliometric tools or journal rankings for reducing complexities is a feature that managers, policy-makers and academics find appealing. Systems of incentives are designed for steering scientists at a distance and indicators of scientific activities become more and more targets for scientists. It is the era of the governance of science by indicators.

There is no time now for reconstructing how the main characteristics of this era emerged. My thesis here is that the adoption of a research policy based on research evaluation systems rewarding performance of scientists could be at the root of many deep problems afflicting contemporary science. And it could be also responsible for the growing difficulties of using the best available science for governmental regulation and policy, in particular for economic policy.

1. Two environmental contexts.

There are two main environmental contexts where governance by indicators is practised. The first is the publish or perish (PoP) general environment where, according to the common wisdom, competition between scientists improve science and performance indicators based on publications and citations are the money in the market for ideas. Journals attribute prestige and then value to a publication whatever its content. In the PoP environment scholars publish for to add new lines to their curriculum with a view to being hired, promoted, tenured or awarded by some committees.

The second is an environment where PoP is reinforced by centralized research assessment systems and by systems of funding based on performance. The most extreme way to realize a system of incentives of this kind is to assign directly cash for scientific publications as in China (Franzoni et al. 2011). But most of the time, the main tool used to steer research institutions and scientists at a distance is the realization of centralized massive research assessment exercises such as REF in UK and VQR in Italy. The results are then used by governments to decide about funding research institutions. Indeed, in Italy the academic review system has become a laboratory for an unprecedented in vivo experiment in governing and controlling research and teaching via automatic bibliometric tools. Metrics are currently used for distributing small research grants to individual scientists, for deciding about qualification of the candidates for the role of professors, and for deciding about eligibility of professors in hiring committees.

Scientists' behaviours change as a response to the systems of incentives they face. Some changes may benefit science and society. Others, often unintended, harm science and society.

Many papers exist about the effect on science of the PoP environment. I will focus on the effect of the adoption of centralized massive research assessment. There are few quantitative systematic studies about the effects of the introduction of research evaluation on scientific performance, scientists' behaviour or the quality of research. And there are also a few qualitative studies on how the research evaluation systems trickles down into working practices and institutions. My point is that academic malpractice/misconduct/bad science might be considered as strategic responses of researchers to research assessment rules adopted by government and agencies.

[1] Department of economics and statistics, University of Siena, Italy. Institute For New Economic Thinking Grant ID # INO17-00015 is gratefully acknowledged. Paper presented at the Global Solutions Summit, Berlin, 29 May 2018.

2. **Known effects of research evaluation systems**

Some effects of research assessment are well known and documented. The first one is the reduction of intellectual diversity and pluralism. According to (Whitley 2007) a strong research evaluation system and a strong cohesive scientific elite reinforce each other in reducing pluralism. Frederic Lee and coauthors documented the progressive elimination of heterodox economics from UK economics, the concentration and homogenization of mainstream economics and the dominance of selected group of economics departments (Lee and Harley 1998; Lee 2007, 2008; Lee et al. 2013). In Australia, in 2011, government decided to change the rule of evaluations after having observed the "ranking [of journals] were being deployed inappropriately … in ways that could produce harmful outcomes" such as suppressing research of mainly local value in favour of research which would have been published in ranked journals.

A second documented problem regards the suppression of interdisciplinary research. That may be induced by the systematic use of journal rankings (Rafols et al. 2011); (Willmott 2011) argued that in business and management the use of journal lists for evaluating research results in "a loss of broader relevance and suffocation of wider social engagement". It is documented also that UK Reseach assessment exercise, based exclusively on peer review, hindered multidisciplinarity in management studies (McNay 1998).

Other known effects are that researchers follow the herd, as stated by Heckman; their goals are displaced for instance by avoiding risk in selecting topics (Wouters et al. 2015) .

3. **Different indicators, different bad behaviours.**

A growing stream of literature highlight the relation between incentives, performance measurement, and risks for scientific integrity. (Edwards and Roy 2017) spoke of "growing perverse incentives in academia" and compiled a table listing for each incentive, the intended effect and the actual bad effect. (Smaldino and McElreath 2016) spoke of a "natural selection of bad science". They construct an evolutionary model of science where incentives for publication quantity drive the emergence and reproduction of poor methodological practices. (Horton 2015) with a special focus on medicine wrote in *Lancet* that

> "The apparent endemicity of bad research behaviour is alarming. We aid and abet the worst behaviours. Our acquiescence to the impact factor fuels an unhealthy competition to win a place in a select few journals. Journals are not the only miscreants. Universities are in a perpetual struggle for money and talent, endpoints that foster reductive metrics, such as high-impact publication. National assessment procedures, such as the Research Excellence Framework, incentivize bad practices. And individual scientists, including their most senior leaders, do little to alter a research culture that occasionally veers close to misconduct." (Horton 2015)

What are the ways in which these bad behaviours manifest themselves? The tip of the iceberg are retractions. A retraction of a scientific publication is a public statement indicating that the original publication should not have been published and that information contained in it should not be used for future research (Todeschini and Baccini 2016). "Retraction is a mechanism for correcting literature and alerting readers to publications that contain such serious flawed or erroneous data that their findings and conclusions cannot be relied upon" (Ethics 2009)

There is a consensus that the "publish or perish" pressure is connected to plagiarism, data fabrication, and other malpractices from which retractions originate. The connection between retractions and massive research evaluation systems may be illustrated for the case of China, where, according to a letter published by *Nature* (Yang 2016) scientists are rewarded by payments and promotions for publishing in the 'top' journals [there is an A-list of journals]. Arguably as a consequence, not only a growing number of papers from China were unsuccessfully submitted to top journals such as *Science* (Franzoni et al. 2011), but, more seriously, China "contributed well over half of the papers retracted for compromised peer review from 2012

to 2016", according to Retraction Watch database (Huang 2017). The most striking case happened in April 2017 when *Tumor Biology,* a journal at that time published by Springer, retracted 107 papers for faked reviews realized by using fabricated e-mail addresses. In July 2017, the Chinese government investigated that episode and found that 486 authors were guilty of misconduct at various levels.[2]

Why do I speak of a tip of the iceberg? Because retractions are visible, they are growing, but they represent only a small fraction of published literature: in 2010 there were 1.6 retraction for a thousand published articles (Grieneisen and Zhang 2012).

Beneath the tip of the iceberg are the results of behaviours that scholars adopt for respecting or gaming the rules of the research evaluation systems. Here there is plenty of scattered evidence. I think that there are two main types of bad practices.

The first kind consists in modifying the contents of the paper for improving indicators of individual productivity. The second kind consists in modifying contents of the paper for improving impact indicators of papers and scholars. Mario Biagioli called that second kind of misconduct "post publication misconduct" (Biagioli 2016). (I am not sure that this is a post-publication misconduct only. Indeed here we have a publication misconduct altering the reference list of a paper in view of modifying the impact of other papers of the same authors or of other academics that the author want to promote for whatever reason.)

The practice of salami slicing is the best known example of the first kind of malpractice, but it is also the most difficult to document in empirical study (Smolčić 2013; Wouters et al. 2015). It's probably the oldest of the modern type of misconducts: it was indeed stigmatized in a comment in *Science* in 1981. It is a form of redundant publication obtained by fragmenting research into many small papers with similar hypothesis, methodology or results. Scholars pursue fragmentation until they reach the minimal publishable unit. (Broad 1981). In medicine, the duplicate publication "may lead to the incorporation of false evidence in meta-analyses, which will adversely affect the assessment of the efficacy and safety of interventions" (Saiz et al. 2018). In a striking ongoing case, 39 papers authored by a group of Italian scholars are under investigation and 10 were already retracted (Saiz et al. 2018).

A second practice used for improving performance indicators consists in modifying the by-line of articles by adding authors who did not participate directly to the research. Also in this case documenting misconducts is a very difficult task. About this there is a diffuse perception that the question of authorship abuse is endemic. In a recent poll conducted by Times Higher Education, two thirds of the self-selecting 364 respondents to the survey "report having felt slighted by a senior academic over an authorship credit", and one third declared that in their experience for being an author is sufficient to be more senior than the people who did the experiments. (Else 2017) According to another survey from 30% of respondents in economics to a bit more than 40% in medicine reported that honorary authors had been added to their papers (Fong and Wilhite 2017). This kind of misconduct is particularly incentivized if an evaluation system adopts a full-counting approach to the calculus of individual productivity. It is the case, for example of systems where the h-index is used as produced by major bibliometric databases (WoS and Scopus).

Consider the case of Italy. In all national evaluation procedures, Italy started using this kind of indicators in 2011.[3] In 2017, one of the members of the agency of evaluation declared that in the aftermath of the research

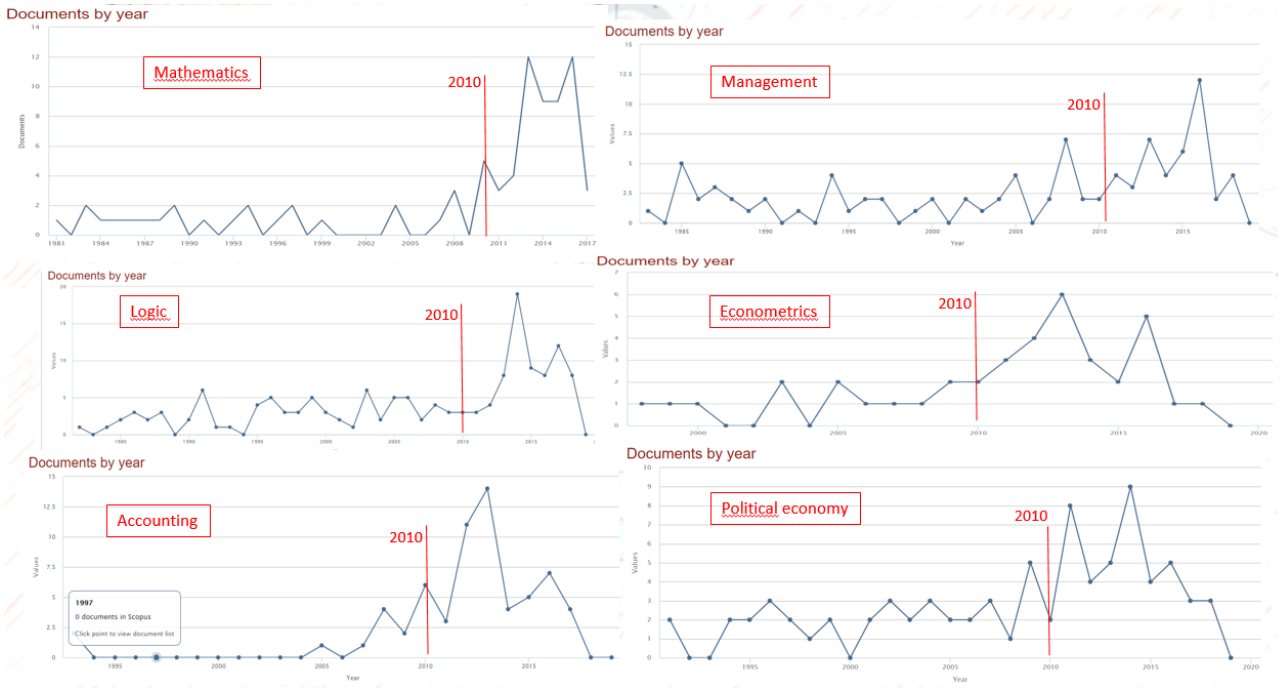
[2] (http://www.xinhuanet.com/english/2017-07/28/c_136480677.htm).

[3] On the technical problems of research evaluation in Italy you can see: Alberto Baccini and Giuseppe De Nicolao, 'A Letter on Ancaiani Et Al. 'Evaluating Scientific Research in Italy: The 2004-10 Research Evaluation Exercise'', *Research Evaluation,* (2017a), Alberto Baccini and Giuseppe De Nicolao, 'Errors and Secret Data in the Italian Research Assessment Exercise. A Comment to a Reply', *RT. A Journal on Research Policy and Evaluation,* 5/1 (2017-07-22 2017b), Sergio Benedetto et al., 'Reply to the Letter on Ancaiani Et Al. 'Evaluating Scientific Research in Italy: The 2004–10 Research Evaluation Exercise'', *Research Evaluation,* (2017), Alberto Baccini and Giuseppe De Nicolao, 'Do They Agree? Bibliometric Evaluation Versus Informed Peer Review in the Italian Research Assessment Exercise', *Scientometrics,*

assessment in his university, the director of a department “wrote to his colleagues to ensure […] that those who were inactive, i.e. with fewer publications than required, were included as co-authors.” (Bonaccorsi 2017). He added also that he “saw with his own eyes the table of a research field in which a list was made of the papers submitted to journals or already accepted, with a scientific division of cases in which the authors (all young) would be asked to add the name of another before the final publication. All that controlled by a well-organized group of professors” (Bonaccorsi 2017).

My guess is that at an aggregate level, these practices results in a probably exceptional growth in the number of papers-per scholar in Italy. In Figure 1 there are the performances of a few senior professors who needed to improve their productivity for reaching the threshold necessary to being admitted to the drawing for being part of the committee deciding on the national qualification of their research sector.

Figure 1. Number of publications per year of some professors selected as members of the committees for the national qualification to the role of professors (ASN 2016). (Source Scopus. May 2018)



Much anecdotal evidence exists about similar patterns of publication for junior researchers who needed to improve their productivity to qualify for the national qualification.

108/3 (2016b), 1651-71, Alberto Baccini and Giuseppe De Nicolao, 'Reply to the Comment of Bertocchi Et Al', *Scientometrics,* 108/3 (2016a), 1675-84, Giovanni Abramo and Ciriaco Andrea D’angelo, 'On Tit for Tat: Franceschini and Maisano Versus Anvur Regarding the Italian Research Assessment Exercise Vqr 2011–2014', *Journal of Informetrics,* 11/3 (8// 2017), 783-87, Giovanni Abramo and Ciriaco Andrea D’angelo, 'Refrain from Adopting the Combination of Citation and Journal Metrics to Grade Publications, as Used in the Italian National Research Assessment Exercise (Vqr 2011–2014)', *Scientometrics,* 109/3 (2016), 2053-65, Giovanni Abramo and Ciriaco Andrea D'angelo, 'The Vqr, Italy's Second National Research Assessment: Methodological Failures and Ranking Distortions', *Journal of the Association for Information Science and Technology,* (2015), n/a-n/a, G. Abramo, T. Cicero, and C. A. D'angelo, 'National Peer-Review Research Assessment Exercises for the Hard Sciences Can Be a Complete Waste of Money: The Italian Case', *Scientometrics,* 95/1 (2013), 311-24, Fiorenzo Franceschini and Domenico Maisano, 'Critical Remarks on the Italian Research Assessment Exercise Vqr 2011–2014', *Journal of Informetrics,* 11/2 (5// 2017), 337-57.

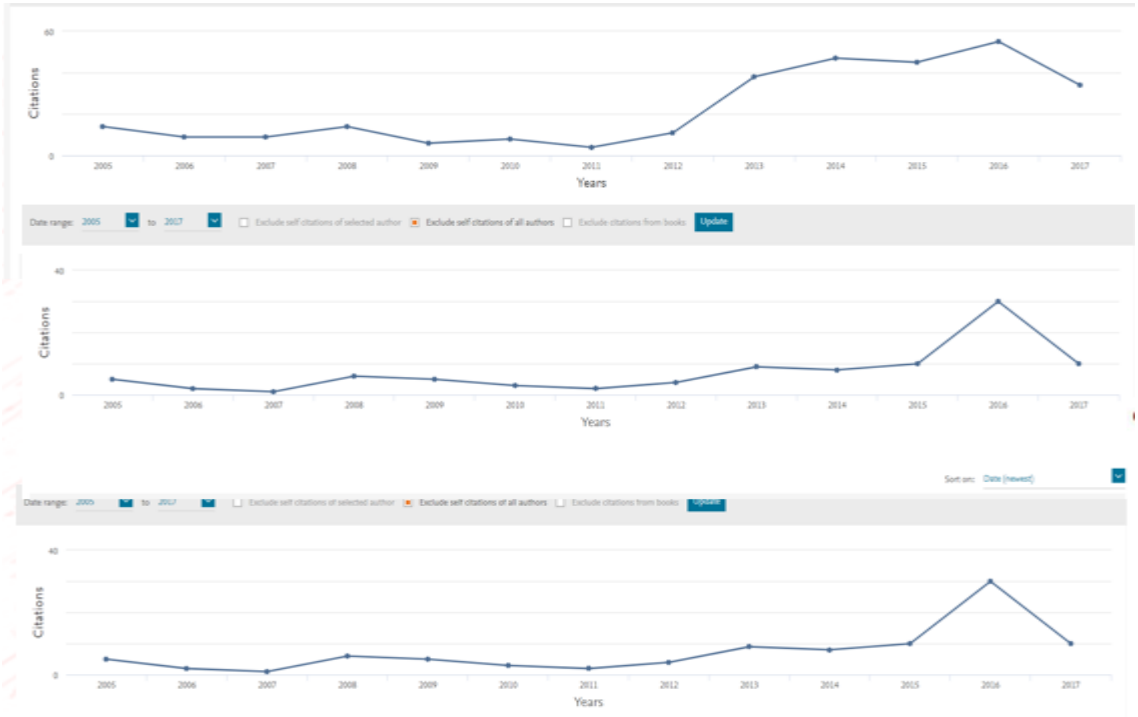
From my point of view, these examples show that scientists publish their articles not for communicating new findings to other scholars, but with a view to insert new lines in their cv to match the performance metrics used by the research evaluation system.

We can consider now the second kind of bad practices aiming to modify citations of articles. Here we have two kinds of misconduct, the first aimed at improving the citation score of an author, the second aimed at altering the citation score of journals for improving the positon of a journal in a ranking.

I use again some Italian examples, since thresholds based on number of citations were introduced as mandatory for the qualification to the roles of professors or for being part of commissions deciding on qualification, and hiring and promotion at a university level.

In Figure 2 you can see some typical profiles (corresponding to some of the profiles of figure 1), where the boom in citations happens after the introduction of the rule.

Figure 2. Number of citations of some professors selected as members of the committees for the national qualification to the role of professors (ASN 2016). (Source Scopus. May 2018)



How is it possible to boost citations? A common practice consists in forming citation rings or citation cartels: a small group of people agrees on exchanging citations. A very simple practice consists in self-citing their previous paper. That last case has been recently documented for a small group of research fields (Seeber et al. 2017).

4. Are research assessment and performance-based research funding corrupting science?

My main point is that research evaluation systems and the adoption of performance-based research funding are not only endangering pluralism in science and in economics, but they are also corrupting science.

Scientific findings acquire their credibility by the “public use of reason”. Since the scientific revolution, scientific truth emerged from a battlefield of conflicting ideas. Now scientific truth appears to be increasingly

guaranteed by labels – that is, by some "objective" ranking of scientists, journals or universities. In research evaluation systems, it's the governments' "objective" ranking of scientists, journals, or universities that guarantees the label of truth to science. Research evaluation systems crowd-out the public use of reason from science, and transform science into a dogma. When it is transformed in dogma, science loses its credibility and may sink, together with political power, in the estimation of citizens of democratic countries.

"Our society will benefit from using the best available science for governmental regulation and policy. One can only applaud when governments want to support the best possible science, invest in it, find ways to reduce biases, and provide incentives that bolster transparency, reproducibility, and the application of best methods to address questions that matter. However, perceived perfection is not a characteristic of science, but of dogma." (Ioannidis 2018)

Moreover, research evaluation systems, by introducing incentives that reward metrics of production and impact corrupt the social norms and values that should regulate interactions in science. Priority in discovery ceases to be the main goal of a scientist. The new goal is gaining appropriate labels and performing better on metrics.

5. Do research evaluation systems put freedom of science at risk?

The risk is that institutions responsible for evaluating research are too close to government and may slowly push science in directions desired by political power, as in planned economies. There is a debate on this point. Research assessments both in UK and Italy have been charged of being forms of managerial Stalinism or Lysenkoism (Amann 2003; Baccini and De Nicolao 2018; Brandist 2014, 2016).[4]

For illustrating the point we can consider the case of economic policy. A government-driven research evaluation may well be designed so as to reward papers and journals that support the economic policies adopted by the government, for example expansionary austerity. The emergence of a mainstream economic policy may be reinforced by stiffening the research evaluation mechanisms, e.g. by ranking mainstream journals in top positions, as happened for example in Italy. [5]

6. A not-so-modest proposal for research policy.

I think it is time to reverse the direction of the last thirty years of science and research policies. The first priority is to minimize the impact of research evaluation systems on science.

A first step in this direction should be the decoupling of evaluation practices from automatic funding mechanisms. As noted by Wouters, this may not be sufficient to ensure a change in actual researchers'

[4] A delicate question concerns the structure of the agency in charge of assessment and its distance from government. In this respect Italy and UK differ. While in the UK the agency in charge of assessment is a quango (quasi autonomous non-governmental agency), acting at a distance from government, in Italy government controls directly the agency in charge of the evaluation ; Roberta Calvano, *La Legge E L'università Pubblica. I Principi Costituzionali E Il Riassetto Dell'università Italiana* (Napoli: Jovene, 2012).

[5] Italy provides a striking example of this kind of problems. In the research assessment exercise for the 2004-2010, the panel for economics was constituted by a small clique of co-authors: 9 panellists out of 20 were directly co-authors (45%). Out of the 8 panellists not affiliated to Italian universities, 5 (63%) were co-authors of at least an Italian panellist. And, last but not least, 9 panellists out of 20 (45%) were among the founders of a micro-ultra-liberal party that participated in the political elections of 2013 by obtaining 0.9% of the votes and no representatives in parliament. This panel developed its own journal ranking, chose reviewers and in many cases directly assigned scores to papers submitted. A. Baccini, 'La Vqr Ed Oltre. Riflessioni Sulla Possibilità Di Valutare Credibilmente La Ricerca in Italia', in Giuseppe Conte (ed.), *Evoluzione E Valutazione Della Ricerca Giuridica* (Napoli: Edizioni Scientifiche Italiane, 2016). Marcella Corsi, Carlo D'Ippoliti and Giulia Zacchia documented the consequences of the decisions of that panel.

priorities if the general landscape continues to be dominated by the performance-informed mood of publish or perish (Wouters 2014).

A second and more effective step should be the dismissal of research evaluation systems and their substitution with less invasive “observatories of research” whose aim, rather than evaluation, should be the description of the research activities conducted by researchers of a country, of universities and so on.

There is a growing bottom-up pressure for replacing the myth of “excellence” with the idea of open and sound science (Moore et al. 2017). To proceed in this direction, government-based regulation and funding incentives should be designed as to reward scholars who produce solid science and provide data, methods, protocols, software openly available for replication and reuse. Only a strong change in research policy will guarantee pluralism; and only pluralism in science enables good policy advices. Good policy advices arise from the battle of ideas in science, not from the battle for achieving labels or conquering a publication spot in a top journal.

References

Abramo, G., Cicero, T., and D'Angelo, C. A. (2013), 'National peer-review research assessment exercises for the hard sciences can be a complete waste of money: The Italian case', *Scientometrics,* 95 (1), 311-24.

Abramo, Giovanni and D'Angelo, Ciriaco Andrea (2015), 'The VQR, Italy's second national research assessment: Methodological failures and ranking distortions', *Journal of the Association for Information Science and Technology*, n/a-n/a.

Abramo, Giovanni and D’Angelo, Ciriaco Andrea (2016), 'Refrain from adopting the combination of citation and journal metrics to grade publications, as used in the Italian national research assessment exercise (VQR 2011–2014)', *Scientometrics,* 109 (3), 2053-65.

--- (2017), 'On tit for tat: Franceschini and Maisano versus ANVUR regarding the Italian research assessment exercise VQR 2011–2014', *Journal of Informetrics,* 11 (3), 783-87.

Amann, Ron (2003), 'A Sovietological View of Modern Britain', *The Political Quarterly,* 74 (4), 468-80.

Baccini, A. (2016), 'La VQR ed oltre. Riflessioni sulla possibilità di valutare credibilmente la ricerca in Italia', in Giuseppe Conte (ed.), *Evoluzione e valutazione della ricerca giuridica* (Napoli: Edizioni Scientifiche Italiane).

Baccini, A. and De Nicolao, G. (2018), 'How Pseudoscientific Rankings Are Distorting Research', *www.ineteconomics.org*.

Baccini, Alberto and De Nicolao, Giuseppe (2016a), 'Reply to the comment of Bertocchi et al', *Scientometrics,* 108 (3), 1675-84.

--- (2016b), 'Do they agree? Bibliometric evaluation versus informed peer review in the Italian research assessment exercise', *Scientometrics,* 108 (3), 1651-71.

--- (2017a), 'A letter on Ancaiani et al. ‘Evaluating scientific research in Italy: the 2004-10 research evaluation exercise’', *Research Evaluation*.

--- (2017b), 'Errors and secret data in the Italian research assessment exercise. A comment to a reply', *RT. A Journal on Research Policy and Evaluation,* 5 (1).

Benedetto, Sergio, et al. (2017), 'Reply to the letter on Ancaiani et al. ‘Evaluating Scientific research in Italy: The 2004–10 research evaluation exercise’', *Research Evaluation*.

Biagioli, Mario (2016), 'Watch out for cheats in citation game', *Nature,* 535, 201.

Bonaccorsi, A (2017), 'La valutazione della ricerca produce effetti indesiderati?', in Angelo Turco (ed.), *Culture della valutazione* (Roma: Carocci), 27-48.

Brandist, Craig (2014), 'A very Stalinist management model', *Times Higher Education*.

--- (2016), 'The risks of Soviet-style managerialism in UK universities', *Times Higher Education*.

Broad, William J. (1981), 'The Publishing Game: Getting More for Less', *Science,* 211 (4487), 1137-39.

Calvano, Roberta (2012), *La legge e l'università pubblica. I principi costituzionali e il riassetto dell'UNiversità italiana* (Napoli: Jovene).

Edwards, Marc A. and Roy, Siddhartha (2017), 'Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition', *Environmental Engineering Science,* 34 (1), 51-61.
Else, Holly (2017), 'Authorship wars: academics outline the rules for recognition', *Times Higher Education*.
Ethics, COPE - Committee On Publication (2009), 'Retraction Guidelines', *http://publicationethics.org/resources/guidelines*, 1-6.
Fong, Eric A. and Wilhite, Allen W. (2017), 'Authorship and citation manipulation in academic research', *PLOS ONE,* 12 (12), e0187394.
Franceschini, Fiorenzo and Maisano, Domenico (2017), 'Critical remarks on the Italian research assessment exercise VQR 2011–2014', *Journal of Informetrics,* 11 (2), 337-57.
Franzoni, Chiara, Scellato, Giuseppe, and Stephan, Paula (2011), 'Changing Incentives to Publish', *Science,* 333 (6043), 702-03.
Grieneisen, Michael L. and Zhang, Minghua (2012), 'A Comprehensive Survey of Retracted Articles from the Scholarly Literature', *PLOS ONE,* 7 (10), e44118.
Horton, Richard (2015), 'Offline: What is medicine's 5 sigma?', *The Lancet,* 385 (9976), 1380.
Huang, Echo (2017), 'China publishes more science research with fabricated peer-review than everyone else put together', *Quartz*.
Ioannidis, John P. A. (2018), 'All science should inform policy and regulation', *PLOS Medicine,* 15 (5), e1002576.
Lee, Frederic S. (2007), 'The Research Assessment Exercise, the state and the dominance of mainstream economics in British universities', *Cambridge Journal of Economics,* 31 (2), 309-25.
--- (2008), 'A case for ranking heterodox journals and departments', *On the Orizon,* 16 (4), 241-51.
Lee, Frederic S. and Harley, Sandra (1998), 'Peer Review, the Research Assessment Exercise and the Demise of Non-Mainstream Economics', *Capital & Class,* 22 (3), 23-51.
Lee, Frederic S., Pham, Xuan, and Gu, Gyun (2013), 'The UK Research Assessment Exercise and the Narrowing of UK Economics', *Cambridge Journal of Economics,* 37 (4), 693-717.
McNay, Ian (1998), 'The Research Assessment Exercise (RAE) and after: 'You never know how it will all turn out'', *Perspectives: Policy and Practice in Higher Education,* 2 (1), 19-22.
Moore, Samuel, et al. (2017), '"Excellence R Us": university research and the fetishisation of excellence', *Palgrave Communications,* 3, 16105.
Rafols, I., et al. (2011), 'How journal rankings can suppress interdisciplinary research. A comparison between Innovation Studies and Business & Management'.
Saiz, Luis Carlos, Erviti, Juan, and Garjón, Javier (2018), 'When authors lie, readers cry and editors sigh', *BMJ Evidence-Based Medicine*.
Seeber, Marco, et al. (2017), 'Self-citations as strategic response to the use of metrics for career decisions', *Research Policy*.
Smaldino, Paul E. and McElreath, Richard (2016), 'The natural selection of bad science', *Royal Society Open Science,* 3 (9).
Smolčić, Vesna Šupak (2013), 'Salami publication: definitions and examples', *Biochemia Medica,* 23 (3), 237-41.
Todeschini, Roberto and Baccini, Alberto (2016), *Handbook of Bibliometric Indicators. Quantitative Tools for Studying and Evaluating Research* (Weinheim (Germany): Wiley-VCH).
Whitley, Richard (2007), 'Changing Governance of the Public Sciences', in Richard Whitley and Jochen Gläser (eds.), *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems* (Dordrecht: Springer Netherlands), 3-27.
Willmott, H. (2011), 'Journal list fetishism and the perversion of scholarship: Reactivity and the ABS list', *Organization,* 18 (4), 429-42.
Wouters, P. (2014), 'The citation: from culture to infrastructure', in B. Cronin and Cassidy R. Sugimoto (eds.), *Beyond Bibliometrics. Harnessing Multidimensional Indicators of Scholarly Impact* (Cambridge, Ma.: MIT Press).
Wouters, P., et al. (2015), 'The Metric Tide: Literature Review (Supplementary Report I to the Independent Review of the Role of Metrics in Research Assessment and Management)', (HEFCE).
Yang, Lihua (2016), 'Open up research evaluation in China', *Nature,* 539, 168.