

# Reinforcing Scholarship of Teaching: Inconsistencies in the Predominant Use of Bibliometrics in Promotion and Tenure

Nasir Abdul Latif Sarwani Henry James Manoj Chakravarty\*

College of Medicine & Medical Sciences, Arabian Gulf University, P.O. Box 22979, Manama, Bahrain

## Abstract

Educational institutions have been evaluating scholarship of frontier research as the key factor for decades to acknowledge academic merit and reward faculty with promotion and tenure. While this has created a substantial degree of pressure towards publication of research, it has tragically de-emphasized the role of teaching and scholarship of reflective inquiry. There is a perceived lack of recognition and reward for teaching in institutions of higher learning, and an almost exclusive predominance of use of bibliometrics for evaluating scholarship of teaching. There is a pressing need to address the imbalance between weightage in teaching and research, the competing activities of clinical work, research and administration. This necessitates the development of a structured pathway to teaching competence, an increase in the support systems that permit its implementation and continued use, a profession-wide change in attitudes towards teaching, and practical opportunities that provide more time for teaching (Doctors As Teachers, 2006). We provide an in-depth review of existing literature on the inconsistencies regarding the evaluation of scholarship and academic merit in relation to promotion and tenure, and the debatable and egregious use of the Impact factor (IF) and citation analysis in context. While identifying key documented inconsistencies related to the exclusive use of bibliometrics of IF and Citation analysis for evaluation of faculty performance in institutions of higher education, we feel that there is urgent need to restore pre-eminence of teaching in universities and the incorporation and application of appropriate metrics for evaluating scholarship of teaching and academic merit, and to reinforce its use as the major determinant for granting promotion and tenure.

**Keywords:** Teaching, Research, Academic merit, Tenure, Promotion, Bibliometrics, Impact factor

## 1. Introduction

Universities generally present their aims and mission with the first priority in teaching, followed by research and community service. They acknowledge that teaching is an important function of universities but timidly disavow that it forms its pre-eminent function. There is concern regarding the tendency within academic administrations to focus predominantly on a journal's Impact Factor (IF) when judging the worth of scientific contributions by researchers, affecting promotions, recruitment, and financial bonuses in some institutions for each published paper. This concern is also due to the fact that some of the most powerful incentives in contemporary science actively encourage, reward and propagate poor research methods and abuse of statistical procedures. This process has been described as 'the natural selection of bad science' to indicate that it requires no conscious strategizing nor cheating on the part of researchers. Instead, it arises from the positive selection of methods and habits that lead to publication (Smaldino McElreath 2016). Teaching is a skill that requires experience and reflection in order to improve and excel. Good teaching is an art, not a science. Students are usually unaware that college Professors are not trained teachers. Ironically, while students think that "all profs do is teach", teaching is where Professors receive the least training, if at all (Fertig 2012). The open secret in academia is that professors are paid to publish, not just to teach and many if not all, consider teaching as a distraction to their primary proclivity towards research. Academic leaders are also known to increasingly pursue careers in academic administration and distance themselves from active teaching.

Incorporation of research into universities and its interconnection with teaching dates back historically to the founder of the modern research university, Wilhelm von Humboldt of the University of Berlin in the 19th century (McNeely 2002). Teaching and research are considered by most authorities as reinforcing each other, the so-called mutual 'enrichment dogma', where professors and students work together on research projects. In this way, universities construct a symbiotic relationship between teaching and research where none really exists naturally (Pocklington & Tupper 2002). Questioning the optimal relationship between research and teaching, it has been found that there is a zero relationship between teaching and research at the individual academic and departmental level where, the ability to be a good teacher and the ability to be a good researcher have been positively correlated, whereas the motivation and time to be a good teacher and a good researcher have been negatively correlated (Hattie and Marsh 2004).

## 2. Method

A systematic search for related articles was carried out from journals listed in SCOPUS, Medline, Social Science Citation Index (Thomson Reuters) and ERIC (best database in education) from the Internet. Journals were selected for inclusion in the study by a combination of formal and informal criteria. Firstly, journals publishing

medically related research with the highest “impact factors” were selected, using the 1992 Institute for Scientific Information (ISI) journal ranking (Garfield, 1993). Over 110 journal articles were identified covering a period beginning from 1930 to 2017. Terms like ‘urge to publish’, ‘pressure to publish’, ‘teaching-research nexus’, ‘article retraction’, ‘bibliometrics’, ‘citation indices’, ‘plagiarism’ and ‘research ethics’ were used to identify full-text articles. Reference lists of relevant articles were also reviewed.

### 3. Discussion

#### 3.1 *The case against Bibliometrics*

For tenure track academics, publishing is a strategic preoccupation. Scientists vie for tenure and funding, and these incentivize frequent publishing of large number of academic papers. Evaluation of university faculty in both undergraduate as well as postgraduate institutions is almost exclusively based on their scholarly contribution through the publication of their research in journals, and the simple and objective methods of research assessment using citation data are broadly interpreted as bibliometrics. Citation statistics are perceived as being more accurate since they substitute simple numbers for complex judgments, and are believed to reduce subjectivity of peer review and thus, IF and Citation analysis set the norm. However, this belief is unfounded. The IF measures the average number of citations of articles published in scientific and social science journals. It is a measure of the frequency with which the average article in a journal has been cited, and is a value calculated annually based on the number of times articles published in a journal are cited in two, or more, of the preceding years. Based on the assumption that IF reflects scientific quality, the IF produces a widespread impression of prestige and reputation, though no experimental data support this hypothesis (Brembs et al. 2013). Citation analysis takes into consideration publication by scientists and scholars for assessing their productivity and impact which has now become an industry standard in education. Bibliometrics was pioneered in the United States of America in the 1950s solely for the purpose for tracing ideas, and the impact of scientific and scholarly work. It was not designed to measure quality in anyway or to measure evaluation of individual scientists or universities or academic institutions (Altbach 2006). The Journal IF, was originally created as a tool to help librarians identify journals to purchase, and not as a measure of the scientific quality of research in an article. Suggestions for use of IF with limitations has been provided by the Academia Publishing House. Out of the eleven items recommended for authors, three significant ones are; a) It is critical to understand that the Journal IF will be a quotient factor only and will not be a quality factor; b) Journal IF will not be related to quality of content and quality of peer review; and c) The IF may be incorrectly applied to evaluate the significance of an individual publication or to evaluate an individual researcher. (Academia Publishing House 2013).

Bibliometrics, while commonly used to evaluate the productivity and scientific output of a researcher, have inherent deficiencies. These deficiencies include the fact that citation distributions within journals are highly skewed, the IF gives no information about individual papers in a journal, and the information is surprisingly vague and can be misleading. For papers, instead of relying on the actual count of citations to compare individual papers, people frequently substitute the IF of the journals in which the papers appear. They believe that higher IFs must mean higher citation counts. But this is often not the case! This is a pervasive misuse of statistics that needs to be challenged whenever and wherever it occurs. (Adler et al. 2008). The properties of the Journal IF are field-specific. It is a composite of multiple, highly diverse article types, including primary research papers and reviews (Seglen 1997); Journal IF can be manipulated or ‘gamed’ by editorial policy (PLoS Medicine Editors 2006), and data used to calculate the Journal IF are neither transparent nor openly available to the public from the proprietary stranglehold (Rossner and Hill 2008). The IF is taken by some administrators as a measure of the typical citation rate for the journal, but as these values suggest, for many journals, it isn’t ‘typical’ at all. Internal research demonstrates how a high journal IF can be the skewed result of many citations of a few papers rather than the average level of the majority, reducing its value as an objective measure of an individual paper (Campbell, 2008). Selection for an academic job guided mainly by citation statistics or papers in high impact journals is blind to the difference between someone who creatively develops a research agenda – and is likely to be doing that in ten years, and someone who grinds out papers in a narrow fashionable field (Werner 2015). In most cases only a few articles contribute to the acquisition of the journal IF, excluding any correlation between an article and the journal in which the article was published. A high percentage (90 %) of the IF of the journal Nature in 2004 was generated by only 25 % of the papers published in that year (Nature Editorial 2005).

Additionally, the lack of timely reporting and publication fundamentally impairs the research enterprise, violates the commitment made by investigators to patients and funders, squanders precious time and resources, and threatens to compromise evidence based clinical decision making (Chen, R. 2016). Randy Schekman, recipient of the Nobel Prize for Physiology and Medicine in 2013 has stated that his laboratory would no longer send research papers to the top-tier journals - Nature, Cell and Science by name. He has claimed that pressure to publish in ‘luxury’ journals has encouraged researchers to cut corners and pursue trendy fields of science instead of doing more important work. Criticizing the use of IF, Schekman said it was ‘toxic influence’ on science that

'introduced a distortion' (The Guardian 2013). The San Francisco Declaration of Research Assessment document initiated by the American Society for Cell Biology (ASCB) by a group of editors and publishers has been endorsed by signatories from over 75 institutions and 150 senior figures in science and scientific publishing, and specifically addresses the problem of evaluating the output of scientific research, highlights the misuse of IF as the central problem in this process, and explicitly disavows the use of IF. The declaration has 18 recommendations — targeted at the different constituencies. The first one establishes its over-riding objective: "Do not use journal-based metrics, such as Journal IFs, as a surrogate measure of the quality of individual research articles, to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions." (DORA 2012). Another available citation index, called the h-index of a researcher, utilizes the number of papers co-authored by the researcher with at least h citations each (Hirsch 2005). The h-index is an author level metrics to measure scientific productivity. The h-index is made to address the broad impact of an individual's work. Despite its claims, the h-index has the following disadvantages. The h-index does not account for the number of authors of a paper, it does not account for the typical number of citations in different fields, it does not consider the context of citations, and it can be manipulated through self-citations (Bayeh 2013). In essence, the h-index is limited by the fact that a researcher could have a high h-index and not be the first/last author of a paper and may not have published a single paper with principal authorship in many years and conversely, one could have a low h-index and each one of those papers may have been heavily cited. The possible fallout of this is that a researcher may have a single but widely cited article and be considered exceptional and superior, while some other researcher with several uncited papers may be considered unimportant or inferior. A collective system of evaluation of scientific work assessed with the various methods discussed above, is nonetheless used to fulfill purposes for which they were not intended. Thus, the system used almost universally for the purposes of hiring, promotion and salary review creates significant discrepancies and is hence questionable. Accepting the limitations of IF and citation analyses, modifications have been suggested. These include Fractional Citation Weighting (Zitt and Small 2008), Aggregated Journal-Journal Citation relations (Chen 2008), and the J-index (Bayeh, 2013) but these methods still present an overall correlation with IF. Despite the many conflicting opinions about IF, it is still applied universally for judging scholarship, and the argument unfortunately is that though IF is not a perfect tool to measure the quality of articles, it has the advantage of already being in existence and is, therefore, a good technique for scientific evaluation (Hoefell 1998).

It is noteworthy to consider a few additional flaws in the use of research statistics, and P-values are just the tip of the iceberg. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true (Ioannidis 2005). A reproducibility check of raw data showed that much of the data used were either wrong or made no sense: most of the usable data were missed by their script, 94% of the abstracts that reported P-values greater than 2 had high correlation and overlap between reported outcomes, and only a minority of P-values corresponded to relevant primary outcomes (Ioannidis 2014). Report of a survey published in the journal *Nature* asked scientists what led to problems in reproducibility and more than 60% of respondents said that each of two factors — pressure to publish and selective reporting, always or often contributed (Baker 2016). According to one report, the number of retracted articles had increased tenfold over the past decade (Van Noorden 2008, Grieneisen 2012), and research fraud was the commonest reason cited for retraction of articles (Balhara 2015). This indicates that research is becoming less pioneering or that the objectivity with which results are produced is decreasing, possibly due to the fact that scientists are responding to incentives (Fanelli 2012). It has recently been reported in the journal *Science*, about a scheme to sell authorship on papers already accepted for publication in indexed journals with high IFs, and for \$26,000 the researcher could buy the co-authorship of a manuscript (Yang 2013).

In regards to research productivity, the most fundamental issue of all is to judge the paper, not the journal. Cases of clear scientific misconduct have received significant media attention, but less flagrantly questionable research practices may be more prevalent and, ultimately more damaging to the medical enterprise (John, et al 2012). Research design in general has a history of shortcomings and yet, misuse of statistical procedures and poor methods have persisted and possibly grown. In fields such as psychology, neuroscience and medicine, practices that increase false discoveries remain not only common, but normative (Enserink 2012). The journal *Nature*, stated in a 2006 editorial, "Scientists understand that peer review per se provides only a minimal assurance of quality, and that the public conception of peer review as a stamp of authentication is far from the truth". It appears therefore, that evaluating research productivity as the hallmark of academic performance and excellence by the predominant use of bibliometrics is flawed at the outset. Having said that, it amounts to an inversion of priorities to put the major emphasis on making evaluation of published papers more objective at the cost of neglecting evaluation of quality teaching.

### *3.2 Importance of teaching, its scholarship and its evaluation.*

This brings us to the issue of the importance of teaching in the context of teacher evaluation. When defined as scholarship, teaching educates, entices and inspires learners. As a scholarly enterprise, teaching originates with

existing knowledge that a teacher possesses. Those who teach must, above all, be inspired, well informed, and possess a body of knowledge that is critical to the art of inquiry and dissemination of integrative and analytical process of learning as applicable to their fields. Teaching, at its best, means not only transmitting knowledge, but transforming and extending it as well. (Boyer 1990). There may be a huge discrepancy, 'between what universities say they do to recognize good teaching, and what the majority of their academic staff perceive they do' (Ramsden and Martin 1996). The top priority must therefore be good and competent learning bred in an environment of good and competent teaching. This has been recommended through medical schools being encouraged to emphasize teaching abilities when recruiting and promoting staff (Abrahamson 1991) and to help existing staff to become better teachers. In its present form of evaluation by use of citation data, quality of both teachers and teaching remain compromised in an environment which nurtures gradual diminution of quality in student-teacher interactions, and teachers going through these modified requirements, are now more motivated towards doing research while simultaneously putting the area of teaching into considerable jeopardy.

It is now clear that promotions and distinctions are well within reach of faculty as long as a certain stipulated number of publications have been documented with a clear de-emphasis on commitment towards scholarship of teaching, thus creating a vicious cycle of obtaining academic and social status, and monetary gain. As a result, selection for senior academic posts in general now depend somewhat exclusively on the number and quality of research publications. This tends to pressurize faculty to conduct a variable degree of unnecessary research of compromised quality. Student ratings of teaching are generally not well accepted in academic circles. On the contrary, student ratings have been repeatedly shown to have a high level of validity, and those complaints about them have been debunked by research (McKeachie 1997).

Promotion criteria represents one of the most tangible indicators of the state of teaching in universities. The Dearing Report found that only 3% of academics thought the system rewarded excellence in teaching. The report documented that academic staff thought payment should reward excellence in teaching but were not seen to do so (NCIHE 1997). A clear directive from the British government through the 'WHITE PAPER' (2003) has stated that 'Institutions must properly reward their best teaching staff; and all those who teach must take their task seriously' (DfES, 2003). When done properly, educational research is every bit as demanding, rigorous, and important to the future of an academic discipline as traditional disciplinary research (Huber and Morreale 2002).

In order to support a relevant and scientific audit of teaching, a rubric called the Teaching Portfolio was created to quantify teaching activities and is now widely used. The teaching portfolio is a personal dossier of documents recording teaching performance that is compiled by a faculty member toward documentation and recognition of the act of teaching. It is a factual description of a professor's major strengths and teaching achievements. The Teaching Portfolio has been chosen as a model because it compiles summative and formative evaluation in a single process, honors teaching as a scholarly activity, and it serves as a practical and efficient way to document teaching and its progressive development with time. The portfolio provides a solid basis for evaluating the faculty member's teaching performance and contributions to education (Seldin 1997). For the individual medical teacher, the initial focus should be on fundamentally defining what quality-teaching is, rather than operationalizing the concept of quality based on whatever current assessment measures exist (Harvey 1995).

#### 4. Conclusion

Though the historical concept of the teacher as a 'philosopher and guide' is presently outdated, the most important undertaking of a university is still its teaching of undergraduate students since strong undergraduate teaching is foundational. It is most unlikely for any university to fulfill its mission in the absence of a strong commitment to undergraduate teaching. Of the many roles of universities, teaching undoubtedly represents their pre-eminent commitment followed by research. Most universities proclaim their commitment to teaching through mission statements, learning and teaching strategies but despite this, public endorsement and universities' dedication to teaching remains non-committal in its objective. Academics working in smaller and teaching intensive undergraduate institutions with heavy teaching and consequently diminished research productivity find it impossible to maintain a competitive research profile (Sharobeam and Howard 2002). Assessing academic merit in a fair and balanced way, constitutes a serious challenge. The main incentive and priority should therefore be to prioritize teaching and not be persuaded that teaching should be sacrificed for research or clinical interests. Teaching should be recognized as a core professional activity, and teaching posts with significant, formal education responsibilities should be recruited primarily on the basis of teaching expertise (BME 2006). Additionally, university research is sometimes plagued by unethical conduct. This includes conscious falsification of results, plagiarism, and suppression of research results for non-scholarly reasons. Conflicts of interest occur when professors have commercial and financial interests in the results of their research (Langford 1991). Whenever a quantitative metric is used as a proxy to assess a social behavior, it becomes open to exploitation and corruption, and this is summarized as, 'when a measure becomes a target, it ceases to be a good measure' (López-Cózar, et al 2014). Regarding evaluation of published papers, there is a pressing need to eliminate the use of journal-based metrics, such as journal IF, in funding, appointment, and

promotion considerations due to significant inconsistencies. There is the need to assess research on its own merits rather than on the basis of where the research is published. As the tyranny of bibliometrics tightens its grip, it has a disastrous effect on potential young researchers and forcing research to fit the mold of high-impact journals weakens it (Werner 2015). Emphasis on where research is published, relying on IF to reward academic work with funding or promotion, is ripping the soul out of academia. Impact factors have so much to answer for (Abbasi 2004).

Scholarly productivity and academic merit can be judged using various formats and not only by use of citation metrics. These include memberships on editorial boards, invitations, awards, grant funding, judgment of fellow scientists, letters to journals, blogs, posts etc. since they all project productivity and an active and inquiring mind. Qualitative as well as quantitative research publications need to be supplemented by teacher research, articles that are expository, hypothetical, deductive that is, any form of intellectual exercise that is beneficial to students, academia and to the world outside of the classrooms. The objective assessment of effective teaching should therefore be taken into account as a valid method for evaluation of teaching in addition to the methods proposed above in relation to research productivity. Just as research can no longer be seen as simply discovering or creating knowledge, and teaching is more than simply transmission of what is already known, there are several different relationships between research and teaching and these vary between disciplinary groups (Colbert 2004). The priority with the uncritical pursuit of research has weakened universities' commitment to training of students and quality of undergraduate education. The state of undergraduate education at research universities is such a crisis, an issue of such magnitude and volatility that universities must galvanize themselves to respond (Boyer commission 1997).

Citation analysis appear to be synonymous with quality, and there is an intuitive predicament that high quality articles are highly cited. The precise interpretation of rankings based on citation statistics needs to be better understood. Research is too important to measure its value with only a single measurement tool as research usually has multiple goals and it is therefore reasonable that its value must be judged by multiple criteria. There are a number of caveats in this context. A single number can never give more than a rough approximation to an academic's multifaceted profile, and many other factors should be considered in combination towards evaluating an individual. Additionally, there can always be exceptions to rules and this needs to be kept in mind, particularly in decisions for promotion and tenure. In scientific standards, a citation should refer to the original or primary sources in which results or observations were reported for the first time. Many people, however, disregard this principle and tend to cite review articles rather than original contributors. As a result, Review Journals (journals that publish only literature reviews) harvest the highest impact factors possible compared to primary-research journals. For example, Nature Reviews Genetics has an impact factor higher than Nature itself eg. in 2012, IF 41 vs. IF 38, respectively (Nature, 2014). Clearly, more research in understanding the similarities and differences of distribution of various indices and their distributions in different fields of science in general and medicine in particular would be of value.

The global mindset of using only research as the means of evaluation of scholarship and academic merit needs to change and institutions must recognize and reward excellence in teaching as a major denominator of scholarship and academic merit. There are numerous strategies that can be used to enhance and fortify the mutual relationship. These are firstly, to conceive a reward system with minimum quality threshold in research and teaching with a de-emphasis on rewarding one or the other, and secondly, re-creation of job conditions to meet the needs of the institution. Selective and predominant use of IF at individual and institutional levels are counter-productive. The IF is also artificially manipulated by authorities. The weightage on IF should be removed or de-emphasized from journal assessment or science metrics. The scientific content of a research paper is what should matter, not an abstract number derived with inherent technical shortcomings of basic scientific ethics. If the IF, or any other ranking system continue to be utilized to classify individuals and institutions, journals will continue to reject authors to the detriment of quality and objectivity. The authors recommend that further efforts need to be made to agree on a uniform evaluation model that could effectively incorporate preeminence of scholarship of teaching and academic merit as against existing evaluation metrics that are questionable to say the least, and heavily weighted towards using the bibliometrics of IF.

*"The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor". (Campbell, 1976)*

The authors have no conflict of interest.

**Acknowledgement:** The authors wish to thank Vengilyn Umali Garcia for her meticulous type-setting and text formatting.

## References

1. Abbasi, K. (2004). Let's Dump Impact Factors. *BMJ*; 329:0-h doi.org/10.1136/bmj.329.7471.0-h

2. Abrahamson, S. (1991). The Dominance of Research in Staff of Medical schools: time for a Change? *The Lancet*: 337, 1586-8. doi:10.1016/0140-6736(91)93275-E
3. *Academia Publishing House*. (2013). Impact Factor: Publishing quality article in an Impact Factor Journals. <https://www.academiapublishing.org/pdf/impact%20factor%20and%20index%20copernicus.pdf>
4. Adler, R., Ewing, J., and Taylor, P. (2008) Citation statistics. *A report from the International Mathematical Union*. <https://www.mathunion.org/publications/report/citationstatistics0>
5. Altbach, P.G., (2006). The Tyranny of Citations. *Inside Higher Ed*. <https://www.insidehighered.com>
6. Baker, M. (2016). Is there a reproducibility crisis. *Nature*. 2016 May 25;533(7604):452-4. doi: 10.1038/533452a.
7. Balhara, Y.P.S., Mishra, A. (2015). A study exploring attributes and nature of the retracted literature on mental disorders. *Indian Journal of Medical Ethics*. 12(1).
8. Bayeh, C.Z. (2013). J-index for ranking Scientists and Researchers. *The General Science Journal*. 2 EGRDI transaction on Education, Lebanon. p 1-6 <https://www.researchgate.net/file.PostFileLoader.html?id...assetKey>
9. Board of Medical Education. (2006). Doctors as Teachers. *BMA*. ISBN: 1-905545-06-1
10. *Boyer commission on Educating Undergraduates in the Research University*, (1997). Rein-venting Undergraduate Education: A Blueprint for America's Research Universities. Princeton, NJ: Carnegie Foundation for the Advancement of Teaching.
11. Brembs, B., Button, K., & Munafò, M. (2013). Deep impact: Unintended consequences of journal rank. *Frontier Human Neuroscience*, 7, 291.
12. Campbell, D.T. (1976). On the conflicts between biological and social evolution and between psychology and moral tradition. *Zygon*. 11(3):167-208
13. Campbell, P. (2008). Escape from the impact factor. *Ethics in Science and Environmental Politics (ESEP)*. Vol. 8: Preprint, 2008. doi: 10.3354/esepp00078
14. Chen, C.M. (2008). Classification of Scientific Networks using Aggregated Journal-Journal Citation relations in the Journal Citation Reports. *Journal of the American Society for Information Science and Technology*. 59: (11). 2296-2304. doi:10.1002/asi.20935
15. Chen, R., Desai, N.R., Ross, J.S., Zhang, W., Chau, K.H., Wayda, B., Murugiah, K., Lu, D.Y., Mittal, A., Krumholz, H.M., Hines, H.H. (2016). Publication and reporting of clinical trial results: cross sectional analysis across academic medical centres. *BMJ* 2016;352:i637. doi: <http://dx.doi.org/10.1136/bmj.i637>
16. Colbert, C.L. (2004). *A Cybernetic systems model of teaching and research production: Impact of disciplinary differences*. Paper presented at Research and Teaching: Closing the Divide? An International Colloquium, Marwell Conference Center, Winchester, 17-19 March. [www.solent.ac.uk/r&tconference](http://www.solent.ac.uk/r&tconference)
17. Editorial, *Nature*. (2005). Not-so-deep impact. *Nature*, 435(7045), 1003–1004.
18. *DfES* (2003). The Future of Higher Education (London, UK; HMSO).
19. Enserink, M. (2012). Final Report on Stapel blames field as a whole. *Science* 338, 1270-1271. doi:10.1126/science.338.6112.1270
20. Fanelli, D. (2012). *Scientometrics*. 90: 891. doi:10.1007/s11192-11-0494-7
21. Fertig, J. (2012). How do Professors learn to teach (or do they?). Commentaries. *The John William Pope Center for Higher Education Policy*. [www.popecenter.org](http://www.popecenter.org)
22. Garfield E. Philadelphia, Pa: Institute for Scientific Information; 1993. *SCI Journal Citation Reports: A Bibliometric Analysis of Science Journals in the ISI Database*
23. Grieneisen, M.L., Zhang, M. (2012). A comprehensive survey of retracted articles from scholarly literature. *PLOS One*. 7(10): e44118. doi:10.1371/journal.pone.0044118
24. Harvey, L. (1995). The Key Issues: The Quality Agenda (Editorial), *Quality in Higher Education*, 1: 5-12.
25. Hattie, J., Marsh, H.W. (2004). One Journey to unravel the relationship between research and teaching. "*Research and Teaching: the closing divide? An International Colloquium*". Marwell Conference Centre, Colden Common, Winchester, Hampshire, 18-19 March. [https://cdn.auckland.ac.nz/assets/education/hattie/docs/relationship-betweenresearch-and-teaching-\(2004\).pdf](https://cdn.auckland.ac.nz/assets/education/hattie/docs/relationship-betweenresearch-and-teaching-(2004).pdf)
26. Hirsch JE (2006). An index to quantify an individual's scientific research output. *National Academy of Sciences* 102(46):16569-16572. doi:10.1073/pnas.0507655102
27. Hoeffel, C. (1998). Journal Impact factors. (Letter). *Allergy* 53(12): 1225.
28. Ioannidis, J.P. (2005). Why most published findings are false. *PLoS Med* 2005. Aug2(8)e: 124
29. Ioannidis, J.P. (2014). Why an estimate of science-wise false discovery rate and application to the top medical literature is false. *Biostatistics*. Jan;15(1):28-36; discussion 39-45. doi:10.1093/biostatistics/kxt036. Epub 2013 Sep 25.
30. John, L.K., Loewenstein, G., Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*. 23: 524-532. doi:10.1177/0956797611430953

31. Lanford, J.W., (1991). Secrecy, Partnership and the Ownership of Knowledge in the University. *Intellectual Property Journal*, 6: 155-169.
32. Lather, P. (1993). Fertile obsession: validity after post-structuralism, *Sociological Quarterly*, 34, pp. 673-693.
33. Lopez-Cozar, E.D., Robinson-Garcia, N., Torres-Salinas, D. (2014). The Google Scholar experiment: how to index false papers and manipulate bibliometric indicators. *Journal for the Association for Information Science and Technology*. 65(3): 446-454. doi:10.1002/asi.23056
34. McKeachie, W.J., (1997). "Student Ratings: The Validity of Use," *Amer. Psychologist*, 52(11), 1218.
35. McNeely, I.F. (2002). The Unity of Teaching and Research: Humboldt's Educational Revolution. *Scholarsbank, University of Oregon*. 32-35.  
<https://scholarsbank.uoregon.edu/xmlui/handle/1794/1456?show=full>
36. Nature (2006). What you can't measure, you can't manage: the need for quantitative indicators in peer review. Quality and value: The true purpose of peer review. Peer Review: Debate. | doi:10.1038/nature05032
37. Nature (2014). Website showing the impact factors of NPG journals:  
[http://www.nature.com/npg\\_/company\\_info/impact\\_factors.html](http://www.nature.com/npg_/company_info/impact_factors.html).
38. *PLOS MEDICINE EDITORS*. (2006).  
<http://journals.plos.org/plosmedicine/issue?id=info%3Adoi%2F10.1371%2Fissue.pmed.v03.i06>
39. Pocklington, T., Tupper, A. (2002). *No Place to Learn: Why universities are not working*. UBC Press. Chapter 6, p 115.
40. Ramsden P., Martin, E. (1996) Recognition of good university teaching: policies from an Australian study, *Studies in Higher Education*, 21(3): 299-315.
41. Rossner, Van Epps, and Hill. (2008). Irreproducible results: A response to Thomson Scientific. *J. Cell Biol.* 180, 254–255. doi:10.1083/jcb.200801036
42. San Francisco Declaration on Research Assessment (DORA), (2012). Putting science into the assessment of research. *Annual Meeting of The American Society for Cell Biology (ASCB)* in San Francisco, California, CA, December 16, 2012.
43. Seglan, P.O., (1997). Why the Impact Factor of Journals should not be used for evaluating Research. *BMJ*. 314, 498-502.
44. Seldin, P., (1997). *The Teaching Portfolio: A Practical Guide to Improved Performance and Promotion/Tenure Decisions*, 2nd ed., Anker Publishing Co., Bolton, Massachusetts, MA.
45. Sharobeam, M.H, Howard, K. 2002. Teaching demands versus research productivity. *Journal of College Science Teaching*. 31: 436
46. *THE GUARDIAN*. (2013). Nobel winner declares boycott of top science journals.<https://www.theguardian.com/science/2013/dec/09/nobel-winner-boycottscience-journals>
47. Van Noorden, R. (2008). Science publishing. The trouble with retraction. *Nature*. 478(7367):26-28. doi:10.1038/478026a
48. Werner, R. (2015). The focus on bibliometrics makes papers less useful. *Nature*. 517, 245. Macmillan Publishers Limited.
49. Yang, W. (2013). Research integrity in China. *Science*, 342(6162):1019.
50. Zitt, M. (2008). Modifying the Journal Impact Factor by Fractional Citation Weighting: The Audience Factor. *Journal of the American Society for Science and Technology*. 59: (11). 1856-1860.