The Weakening Relationship Between the Impact Factor and Papers' Citations in the Digital Age

George A. Lozano
Observatoire des Sciences et des Technologies (OST), Centre Interuniversitaire de Recherche sur la Science et la Technologie (CIRST), Université du Québec à Montréal, CP 8888, Succ. Centre-Ville, Montréal, QC. H3C 3P8, Canada and Estonian Centre of Evolutionary Ecology, 15 Tähe Street, Tartu, 50108, Estonia. E-mail: dr.george.lozano@gmail.com

Vincent Larivière
École de bibliothéconomie et des sciences de l'information, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, QC, H3C 3J7, Canada, and Observatoire des Sciences et des Technologies (OST), Centre Interuniversitaire de Recherche sur la Science et la Technologie (CIRST), Université du Québec à Montréal, CP 8888, Succ. Centre-Ville, Montréal, QC. H3C 3P8, Canada. E-mail: vincent.lariviere@umontreal.ca

Yves Gingras
Observatoire des Sciences et des Technologies (OST), Centre Interuniversitaire de Recherche sur la Science et la Technologie (CIRST), Université du Québec à Montréal, CP 8888, Succ. Centre-Ville, Montréal, QC, H3C 3P8, Canada. E-mail: gingras.yves@uqam.ca

Historically, papers have been physically bound to the journal in which they were published; but in the digital age papers are available individually, no longer tied to their respective journals. Hence, papers now can be read and cited based on their own merits, independently of the journal's physical availability, reputation, or impact factor (IF). We compare the strength of the relationship between journals' IFs and the actual citations received by their respective papers from 1902 to 2009. Throughout most of the 20th century, papers' citation rates were increasingly linked to their respective journals' IFs. However, since 1990, the advent of the digital age, the relation between IFs and paper citations has been weakening. This began first in physics, a field that was quick to make the transition into the electronic domain. Furthermore, since 1990 the overall proportion of highly cited papers coming from highly cited journals has been decreasing and, of these highly cited papers, the proportion not coming from highly cited journals has been increasing. Should this pattern continue, it might bring an end to the use of the IF as a way to evaluate the quality of journals, papers, and researchers.

Introduction

The impact factor (IF) was originally devised in the 1960s to guide academic libraries in their journal purchases (Archambault & Larivière, 2009). Although several other types of citation-based measures of journal impact have been recently devised, such as the SCImago Journal Rank (González-Pereira, Guerrero-Bote, & Moya-Anegón, 2009), the Eigenfactor (West & Bergstrom, 2010), and the source normalized impact per paper (Moed, 2010), the two-year IF compiled by Thomson Reuters is still the most widely used. The IF of a given journal for a given year is defined as the mean citation rate, during that given year, of the papers published in that journal during the previous two years. For example, a journal's IF for 2011 considers citations received in 2011 by papers published in that journal during the years 2009 and 2010. Thus, the citation window for individual papers ranges from one year to almost three years, with an average of two years.

Over the past few decades IFs have slowly permeated the collective consciousness of scientists, and IFs have become self-reinforcing measures of journal quality, the papers therein, and their authors. Researchers now consider IFs when choosing their publication outlets; journal editors formulate policies explicitly designed to improve their IFs; and

publishers advertise their IFs on their websites. IFs are often used as a surrogate for the actual number of citations a paper recently published might eventually receive. Such a proxy might be partially justified given that, independently of the quality of the paper, a journal's IF is positively linked with the citations received by its papers (Larivière & Gingras, 2010). Since the early 1990s, as citation data became electronically available, interest and use of the IF has increased, and scholarly articles on the IF have increased exponentially (Archambault & Larivière, 2009).

The digital age also brought forth another change. Since the creation in 1665 of the *Journal des Sçavans* and the *Philosophical Transactions of the Royal Society*, the first two scientific periodicals, researchers have mostly read actual printed journals, so papers published in high-profile journals with high circulation had a greater chance of being read and cited than papers published in less widely available journals. Now that scientific information is disseminated electronically, researchers are less likely to read entire journals; instead, they conduct electronic literature searches on particular topics and find specific articles from a wide variety of journals. Hence, as long as the journal is listed in the main databases (e.g., Web of Science, Scopus, or Google Scholar) and papers are available online, they can be read and cited based on their own merits, unaffected by their journals' physical availability, reputation, or IF.

Therefore, before the digital age the citation rate of any given paper and its journal's IF mutually reinforced each other. A journal's IF was (and still is) based on its individual papers' citation rates, and the citation rate of any individual paper was affected by its journal's circulation and availability, which depended on its IF. Now the former is still true, but if new practices of literature search and usage limit the effect of journal IF on paper citation rates, the correlation between paper citation rate and IF should be decreasing. Additionally, the proportion of highly cited papers coming from the highest IF journals should also be diminishing. Here we examine whether this is indeed the case, and consider the implications to the continued use of the IF on the future of scientific publishing. Data for three groups of disciplines are presented: natural and medical sciences taken together, physics, and social sciences, from 1900 to 2011.

Methods

We used Web of Science (WoS) data from Thomson Reuters from 1900 to 2011, covering all areas of natural sciences, medical sciences, and social sciences. The data set covers the *Century of Science* and *Century of Social Sciences* data sets from 1900–1944, and the Science Citation Index (Expanded), the Social Sciences Citation Index, and the Arts and Humanities Citation Index from 1945–2011. The disciplinary classification of journals used for natural and medical sciences in general, and of physics and social sciences in particular, is an adaptation of the classification used by the U.S. National Science Foundation (NSF), which categorizes each journal in only one discipline and specialty.

The data set included 25,569,603 natural and medical sciences papers, 3,211,026 physics papers, and 879,494 social sciences papers. The number of cited references analyzed was 819,369,970. Humanities papers were excluded from the analysis because of their long citation windows and high uncitedness rates (Larivière, Gingras, & Archambault, 2009), but citations from humanities journals were included. To be included in the analysis, papers had to be published in a journal for which an IF could be calculated. However, references made by excluded papers were considered as citations for other papers included in the analysis. Some journals in some years did not have an IF, either because their papers did not receive any citations during the two-year citation window or because two years must elapse before new journals receive their first yearly IFs.

Given that Thomson Reuters does not compile IFs for the entire period studied, and that the exact method by which it calculates IFs is not entirely clear, and hence irreproducible (Moed & Van Leeuwen, 1995; Rossner, Van Epps, & Hill, 2007), the IF of each journal covered in the database was recalculated. In the Thomson Reuters IF, some types of publications are used to count citations (the nominator), but do not themselves count as "papers" (the denominator). Here, IF was calculated the same way as the Thomson Reuters IF, except that this asymmetry between the numerator and the denominator was corrected. Citations to individual papers were counted during the entire two years following their respective publication year. Hence, IF data were not available for the first two years, and full two-year citation windows were not possible for the last two years, leaving a complete data set of both IFs and citation rates from 1902 to 2009. A large proportion of papers remained completely uncited at the beginning of the period (Larivière et al., 2009); to reduce their weight in the IF-citations relationship, the analyses were also conducted excluding uncited papers, in the calculation of both citation rates and IFs.

Two indicators were used to measure the strength of the relationship between the IF of journals and citations of papers. The first indicator was the coefficient of determination (r^2). Each paper was assigned the IF of the journal in which it was published and the citations it received during two years following its publication year, and the r^2 between the two series of variables was calculated for each year. The second indicator is the yearly percentage of all papers that are both among the most cited papers and published in the most cited journals, and, of the most cited papers, the yearly percentage not published in the most cited journals.

Results

Figures 1–3 present the r^2 between IF and paper citations from 1902–2009, for all disciplines of the natural sciences and medical sciences together (Figure 1), physics (Figure 2), and social sciences (Figure 3). For descriptive purposes, in cases where there was a clear break, these
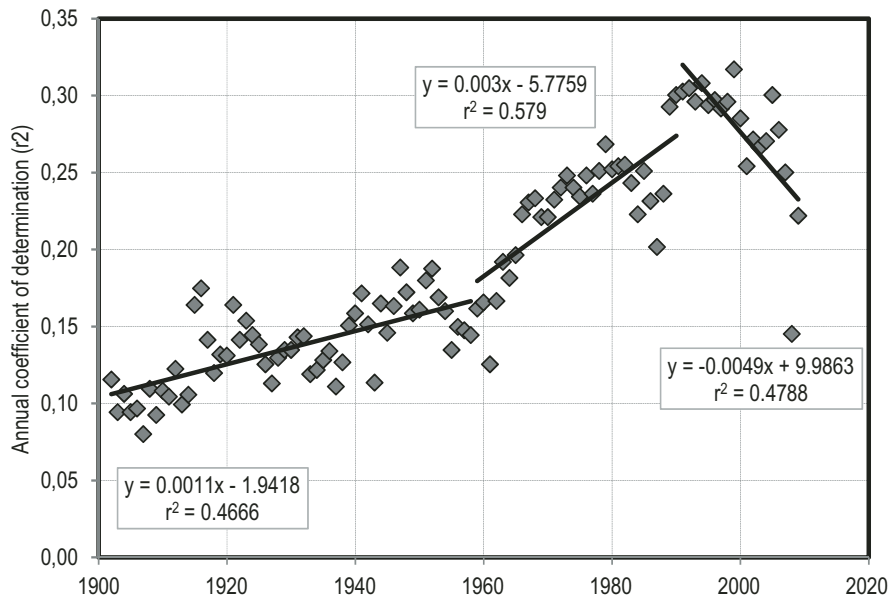
FIG. 1. Coefficient of determination (r^2) between the impact factor of journals and the two-year citation rate of their papers from 1902 to 2009, for all natural and medical sciences journals.
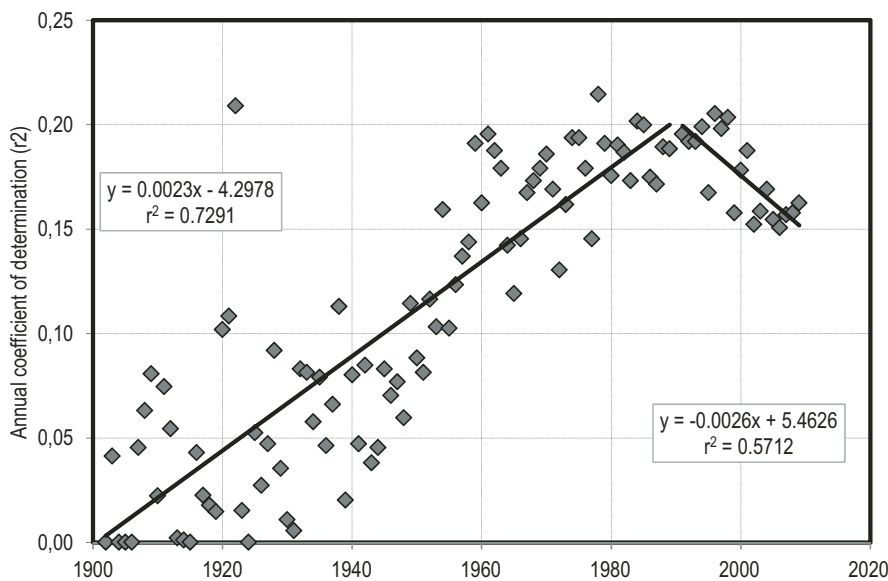

FIG. 2. Coefficient of determination (r^2) between the impact factor of physics journals and the two-year citation rate of their papers from 1902 to 2009.

data were split into 1902–1958, 1959–1990, and 1991–2009. For medical and natural sciences (Figure 1), there was an increase of the correlation between IF and paper citation rates from 1902 until the end of the 1990s. The strength of the relationship between IF and citations did not increase steadily throughout the 20th century. Two dips occurred after the two World Wars, likely as a result of changes in the research system. More interestingly and in contrast to the general pattern throughout most of the 20th century, since scientific information began to be disseminated electroni-

cally, around 1990, the relationship between the IF and citation rates has been weakening.

The same analyses were carried out with two disciplines thought to be at opposite ends of the spectrum of how quickly they made the transition into the electronic domain: physics (Figure 2) and social sciences (Figure 3). Given the smaller sample size, the variation of the r^2 values between IF and papers' citation rates is larger, but in both cases there was a decrease during the last two decades. Although the decrease is not significantly different in the two disciplines,
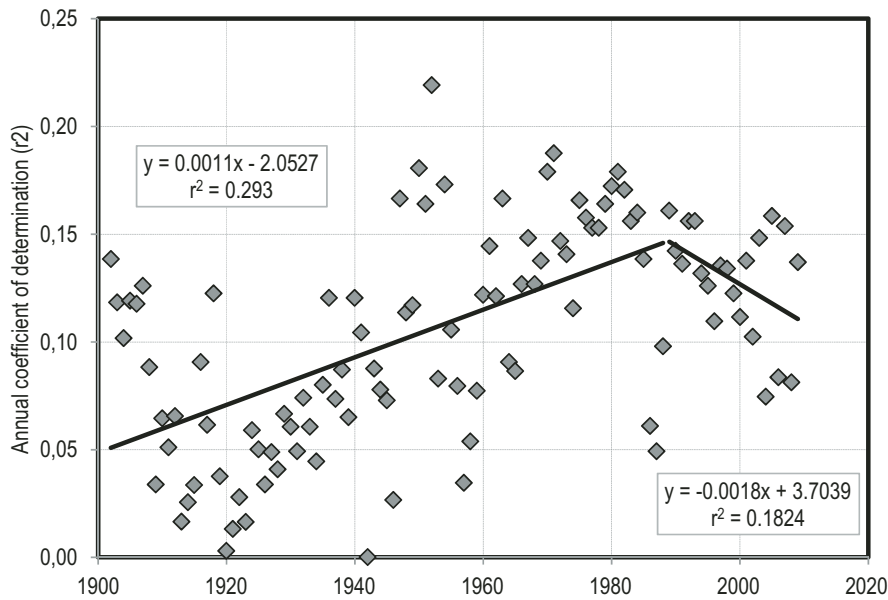
FIG. 3. Coefficient of determination (r^2) between the impact factor of social sciences journals and the two-year citation rate of their papers from 1902 to 2009.
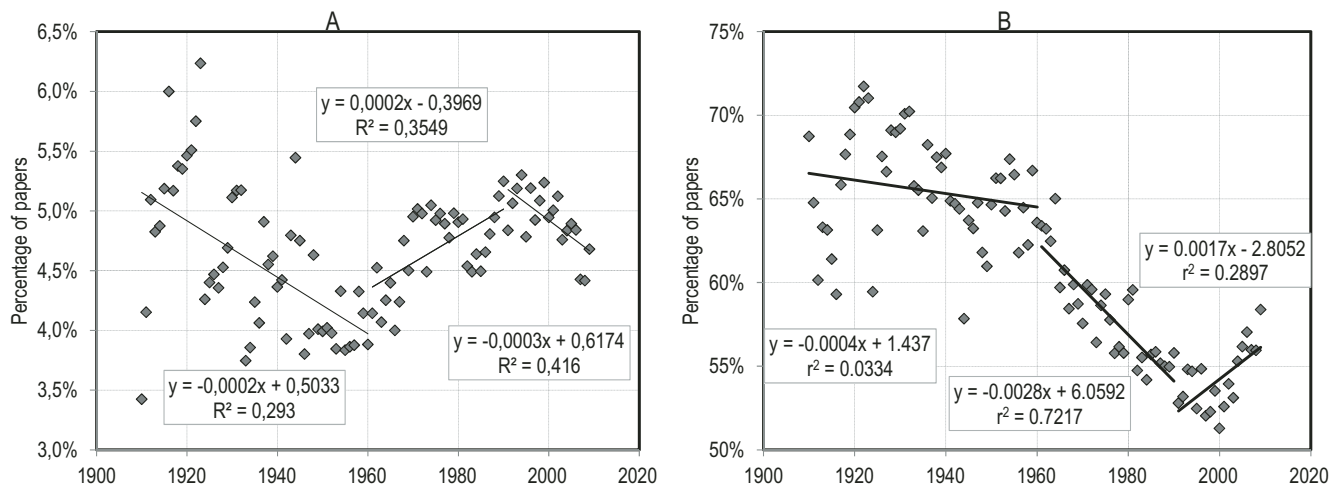

FIG. 4. (A) Of all papers, percentage of the top 10% most cited papers published in the top 10% most cited journals. (B) Of these 10% most cited papers, percentage that were not published in the top 10% most cited journals.

in physics it appears to start earlier, toward the end of the 1980s (Figure 2).

Although not shown, all results are similar and conclusions the same using Pearson's r and Spearman's rank correlation as indicators. All analyses were also carried out excluding uncited papers, both at the level of papers and in the calculation of the journal IFs. When uncited papers are excluded, a clearer trend with fewer fluctuations emerges, but the strength of the relationship between IF and citations remains within the same order of magnitude. So, removing uncited papers does not result in a stronger relationship between the IF and citations.

Figures 4 and 5 present an additional indicator of the relationship between IF and paper citations for all disci-

plines in the natural and medical sciences: of all papers, the percentage of papers that are both in the top 10% (and 5%) most cited and published in the top 10% (and 5%) highest IF journals (Figures 4A, 5A). By contrast, of these top 10% (and 5%) most cited papers, panels B show the percentage that are not published in the top 10% (and 5%) highest IF journals. Both figures show that the relationship between IF and citations has been weakening steadily since 1990, as a larger proportion of top (5% and 10%) most cited papers are published outside journals with top (5% and 10%) IF.

More specifically, of all papers, the percentage of the 10% most cited papers published in the 10% most cited journals has been decreasing since 1990 (Figure 4A), from about 5.25% to 4.50%. Accordingly, of these 10% most
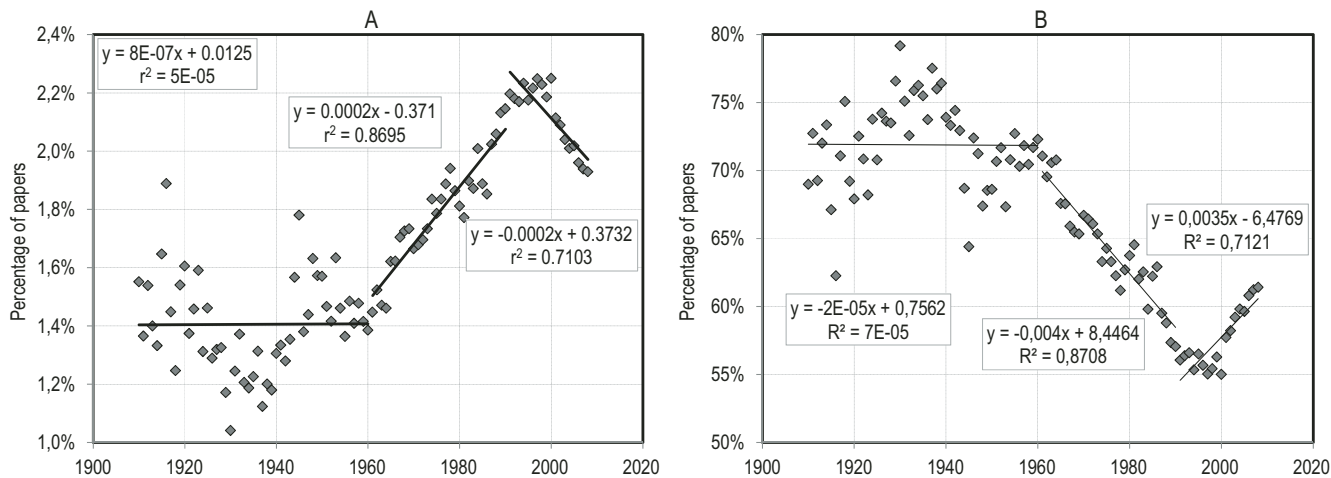
FIG. 5. (A) Of all papers, percentage of the top 5% most cited papers published in the top 5% most cited journals. (B) Of these top 5% most cited papers, percentage that were not published in the top 5% most cited journals.

cited papers, the percentage not published in the 10% journals with the highest IF has been increasing since 1990 (Figure 4B), from 52% to about 56%. This pattern is even more clearly evident when the same comparisons are made for the top 5% of papers and the top 5% of journals (Figure 5). In 1990 about 2.25% of the top 5% papers were published in the top 5% journals, but by 2009 this figure had fallen to 1.90% (Figure 5A). Similarly, in 1990, of the top 5% most cited papers, about 55% were not published in the top 5% journals, but by 2009 the figure had increased to 62% (Figure 5B). Hence, the most important literature is increasingly coming from a greater range of journals, not only the journals with the highest IF.

Discussion and Conclusions

IFs were initially developed to assist libraries in their purchasing decisions, and hence they have had a significant effect on journal circulation and availability. Here we show that throughout most of the 20th century the link between the IF and papers' citations was growing stronger, but, as predicted, this link has been weakening steadily since the beginning of the digital age. This change seems to have started earliest in physics, a field that was quicker to adopt electronic dissemination of information. Also during this time, the percentage of top papers coming from the top journals has been decreasing. Compounded with the fact that, in general, citations have become more widely spread among journals (Larivière et al., 2009), the digital age and its new modes of disseminating and accessing scientific literature might bring forth the end of the IF as a useful measure of the quality of journals, papers, and researchers and have interesting implications for the future of scientific literature.

The IF has been repeatedly criticized for being an unsuitable measure of journal quality (Aksnes, 2003; Morgan & Janca, 2000; Rossner et al., 2007; Rothenberg, 2008;

Seglen, 1997; Vakil, 2005; Whitehouse, 2001). The strongest arguments against its validity and use are (a) some types of publications within journals, such as letters and commentaries, are used to count citations (the nominator), but do not themselves count as "papers" (the denominator), and hence inflate the journal's IF; (b) the IF depends on the number of references, which differs among disciplines and journals; (c) the inclusion of journals in the database depends solely on Thomson Reuters, a private company, and not on scholars and researchers; (d) the exact IF published by Thomson Reuters cannot be replicated using publicly available data; (e) the distribution of citations/paper is not normal, so at the very least the mode or median ought to be used instead of the mean; (f) the 2-year span for papers followed by 1 year for citations is completely arbitrary, and favors high-turnover over long-lasting contributions; and (g) journal editors can manipulate and artificially inflate their IFs. Our analysis identifies one more problem: The relationship between paper quality and IF is weakening, so the IF is losing its significance as a measure of journal quality.

Second, IFs are used as a proxy for paper quality. Except for the most recently published papers that have not had a chance to be cited yet, there is no reason to use the IF as a proxy for a paper's quality. One can readily have access to any individual paper's citation rate and determine how the paper stands on its own, regardless of its journal's IF. As the relationship between paper citation rates and IF continues to weaken, and as more important papers increasingly appear in more diverse venues, it will become even less justifiable to automatically transfer a journal's reputation and symbolic capital onto even its most recently published papers. This should force a return to direct assessments of paper quality, by actually reading them.

Third, and even more troubling, is the three-step approach of using the IF to infer journal quality, extend it to the papers therein, and then use it to evaluate researchers. Our data show that the high IF journals are losing their

stronghold as the sole repositories of high-quality papers, so there is no legitimate basis for extending the IF of a journal to its papers, much less to individual researchers. Moreover, given that researchers can be evaluated using a variety of other criteria and bibliometric indicators (e.g., Averch, 1989; Leydesdorff & Bornmann, 2011; Lozano, 2010; Lundberg, 2007; Põder, 2010), evaluating researchers by simply looking at the IFs of the journals in which they publish is both naive and uninformative.

For the past few centuries journals were a convenient way to organize papers by subject, but search engines now allow us to find individual papers on specific topics from across the entire spectrum of journals, so highly subject-specific journals might become obsolete or begin to amalgamate. Online, open-access journals, such as in the PLoS family of journals, and online databases, such as the ArXiv system and its cognates, will continue to gain prominence. Using these open-access repositories, experts can find publications in their respective fields and decide which ones are worth reading and citing, regardless of the journal. Should the relationship between IFs and papers' citations continue to weaken, the IF will slowly lose its legitimacy as an indicator of the quality of journals, papers, and researchers.

Acknowledgments

References

Aksnes, D.W. (2003). Characteristics of highly cited papers. Research Evaluation, 12(3), 159–170.

Archambault, É., & Larivière, V. (2009). History of the journal impact factor: Contingencies and consequences. Scientometrics, 79(3), 635–649.

Averch, H.A. (1989). Exploring the cost-efficiency of basic research funding in chemistry. Research Policy, 18(3), 165–172.

González-Pereira, B., Guerrero-Bote, V.P., & Moya-Anegón, F. (2009). The SJR indicator: A new indicator of journals' scientific prestige. Available at: http://arxiv.org/ftp/arxiv/papers/0912/0912.4141.pdf

Larivière, V., Gingras, Y., & Archambault, É. (2009). The decline in the concentration of citations, 1900–2007. Journal of the American Society for Information Science and Technology, 60(4), 858–862.

Larivière, V., & Gingras, Y. (2010). The impact factor's Matthew effect: A natural experiment in bibliometrics. Journal of the American Society for Information Science and Technology, 61(2), 424–427.

Leydesdorff, L., & Bornmann, L. (2011). Integrated impact indicators compared with impact factors: An alternative research design with policy implications. Journal of the American Society for Information Science and Technology, 62(11), 2133–2146.

Lozano, G.A. (2010). A new criterion for allocating research funds: "impact per dollar." Current Science, 99(9), 1187–1188.

Lundberg, R. (2007). Lifting the crown-citation z-score. Journal of Infometrics, 1(2), 145–154.

Moed, H.F. (2010). Measuring contextual citation impact of scientific journals. Journal of Informetrics, 4, 265–277.

Moed, H.F., & Van Leeuwen, T.N. (1995). Improving the accuracy of Institute for Scientific Information journal impact factors. Journal of the American Society for Information Science, 46(6), 461–467.

Morgan, V., & Janca, A. (2000). Revisiting the journal impact factor. Australasian Psychiatry, 8(3), 230–235.

Põder, E. (2010). Let's correct that small mistake. Journal of the American Society for Information Science and Technology, 61(12), 2593–2594.

Rossner, M., Van Epps, H., & Hill, E. (2007). Show me the data. The Journal of Cell Biology, 179(6), 1091–1092.

Rothenberg, R. (2008). The impact factor follies. Epidemiology, 19(3), 372.

Seglen, P.O. (1997). Why the impact factor of journals should not be used for evaluating research. British Medical Journal, 314(7079), 498–502.

Vakil, N. (2005). The journal impact factor: Judging a book by its cover. American Journal of Gastroenterology, 100(11), 2436–2437.

West, J.D., & Bergstrom, C.T. (2010). The Eigenfactor metrics: A network approach to assessing scholarly journals. College and Research Libraries Journal, 71, 236–244.

Whitehouse, G.H. (2001). Citation rates and impact factors: Should they matter? British Journal of Radiology, 74(877), 1–3.