

Editorial: Is the Impact Factor (IF) Ethical To Use for Promotion and Tenure Decisions?

Stephen M. Marson, Ph.D., Editor

Journal of Social Work Values and Ethics, Volume 17, Number 1 (2020)

Copyright 2020, ASWB

This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of ASWB.

As editor of the *Journal of Social Work Values and Ethics*, I am committed to enhancing the quality, application and overall use of our journal. There are a variety of paths I pursue to address these goals. One of them is to examine Impact Factor (IF). When I became aware of IF, I thought that acquiring an IF coefficient would be critical for JSWVE. I became particularly startled when a colleague pursuing a full professorship was informed that the Rank and Tenure Committee wanted to assess the IF coefficient for the journals where she had published. In fact, she was discouraged from publishing in any journal that did not have an IF. I became even more startled when I sought a pathway to secure an IF for JSWVE. Once again, I was flabbergasted. There was a fee of \$500 for the journal to obtain an IF coefficient. At that point, I began to wonder if the IF was, in fact, a racket. It was clear that I had to acquire an in-depth knowledge of the IF.

The first task I had to accomplish was understanding the math formulated by Eugene Garfield and Irving H. Sher (Garfield, 2006) that undergirds the concept of IF:

$$A/B = \text{IF (Impact Factor)}$$

A = the number of articles published in given time period that were cited by indexed journals during same that time period.

B = the total number of "citable items" published in the time period found in A.

Everyone who sees this formula will immediately notice that it excludes *citations found in textbooks*. When an author of a textbook cites a journal article, the citation will have a great impact on the largest number of readers. This fact alone greatly challenges the ethics of employing IF for rank and tenure

decisions. However, the IF coefficient is supposed to be a measure of journal usage not of the scholarship of authors who publish in that journal. The logic is, if an author publishes in an often-cited journal, the author's work will be read and used. BUT there is nothing in the formula that suggests that the author's work will actually be read or used merely because it appears in a particular journal! There are at least two fundamental flaws in employing the IF coefficient to assess the scholarly impact of an individual. Overall the IF coefficient includes the negative impact and fails to include some types of positive impact. Both flaws required elaboration.

Negative Impact

Years ago, there was an article that employed the t-test. Statisticians went crazy! The underlying assumptions of the t-test were violated making the overall findings in the publication dubious at best and useless at worst. Among statisticians, the article was cited frequently as an example of how *not* to apply the t-test. In fact, the article was used in a statistics course I took while attending Ohio State where the statistics professor used the article as an example of incompetent employment of a statistic *and* an acknowledgment that the anonymous referee process is not infallible. BAD research is frequently cited! Such publications are a poor reflection on the anonymous referee process for the journal that published the article. Here is the point: If an article is frequently cited because it is severely flawed, the IF coefficient projects the image that the journal has high standards when the *exact opposite* is true. IF is promoted as a measure of the quality of a journal. This is a dubious assumption.

My story of the bad t-test is not an isolated incident. In the 1980s Ned Feder and Walter Stewart initiated a new trend of citing unethical research. If one examines their citations over the decades, one will discover that they are citing journal articles that are so fundamentally flawed scientists agree that the articles should have never been published. In their famous case study of John Darcy, a corrupt researcher from Harvard, Feder and Stewart found more than 200 papers that constituted falsified research. Because Darcy published false research, his findings were profound, incredible and, most importantly, cited elsewhere. When Darcy was uncovered as a fraud, his work stopped being cited as legitimate research. Nevertheless, his printed publications remain housed in libraries internationally. Critically, Darcy's publications are still being cited but as examples of falsified research. Most importantly, the falsified publications of Darcy and others are calculated as part of an IF coefficient. The undisputed fact is: Weak, misleading and false research publications are currently being cited. Bottom line: Journal articles in which the scientific community agree should have never been published are, in fact, included in the IF coefficient. Just because a work is cited many times, it does not mean that the work is worthy of publication. Thus, flaws in the anonymous review process can inflate the IF coefficient.

Positive Impact

In addition to the serious problem of the IF formula including journal articles that should not be included, the IF formula fails to include publications that should be included or have a strong impact on the world of science. We have already noted that articles that have been cited in textbooks are excluded from data used in the IF formula. However, consider the influence of books in general. For example, Jeffrey Alexander's books on theory have had a profound impact on sociologists around the world. Yet, the impact of his incredible contribution is discounted in the IF coefficient. As I write this, I can think of numerous scholarly books that have profound impact on my thought processes. In fact, I was going

to include a list, but the list was too long to publish! None of these profoundly important scholarly contributions are included in the IF coefficient.

One of the greatest impacts a publication can have is when a work of scholarship is required reading for students. Every day, thousands of students are required to read a particular article on closed reserve of the library. The classic article "The Body Ritual among the Nacirema" from 1956 has been read by millions of English-speaking college students across the world. The work produced personal paradigm shifts in the mind of the reader. Hundreds of articles are being read by students. None of these scholarly works are included as part the IF coefficient.

The Application

In my personal search for the impact I have made, I completed a "Google Scholar" search and uncovered an unpublished paper that cited my research on the best methods to teach statistics. As it turns out, my research was used as the conceptual framework for a social work department in their teaching of statistics. In my mind's eye, *that is an impact* that would influence the decision-making process of a Rank and Tenure Committee. Nevertheless, since this unpublished accreditation report is not part of a journal article, the data is not included in the IF coefficient. My research had an obvious impact, but it is not included in the IF coefficient.

Published articles in the *Journal of Social Work Values and Ethics* are highly specialized and will be cited in articles and books written by social workers who have a special interest in social work values and ethics. Of all scholars that currently exist in the world, what is the proportion of those who fit into this highly selective interest group? The answer is somewhere around .05%. Highly specialized journals in any discipline will have a low IF coefficient. Nevertheless, the IF coefficient fails to accurately measure the quality of the article/journal or even its influence on the small proportion of highly specialized readers. Just as critical, clinicians who read an article are not likely to cite it but may adopt the content of an article into their

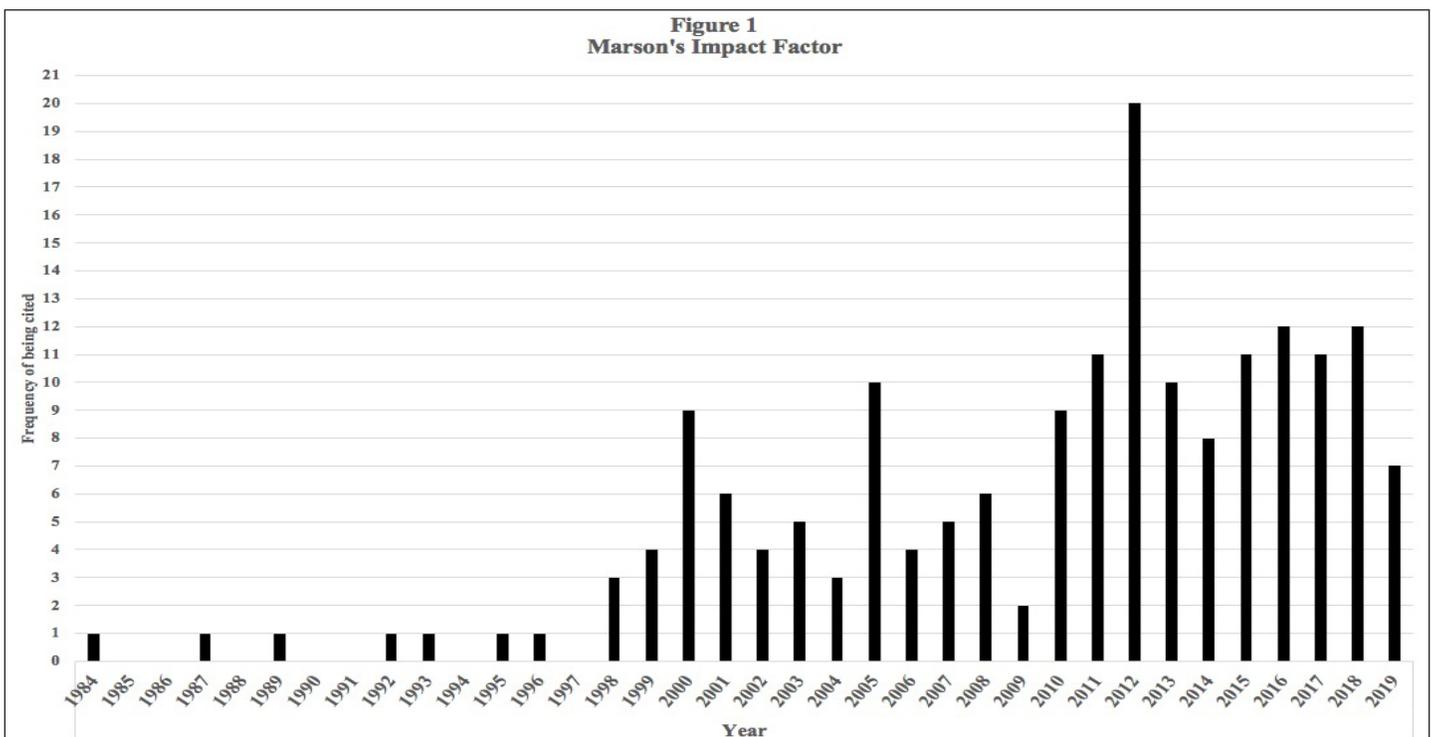
practice model. Such usage is a significant impact in counseling, medicine, psychology, and social work. Yet, it is not included within the IF coefficient.

Hoeffel (1998) notes that the “Impact Factor is not a perfect tool to measure the quality of articles but there is nothing better and it has the advantage of already being in existence and is, therefore, a good technique for scientific evaluation” (p. 1225). Hoeffel’s statement may have been true 21 years ago with the unsophisticated technology of the time, but it is no longer true. Today, as a measure of scholarly influence, IF is *fundamentally flawed*; I believe it is unethical to employ it as an assessment tool for rank and tenure decisions. Let me offer an alternative that is a more accurate portrayal of scholarly impact. Today, the scholar’s resume typically includes a bibliography of references where the author has been cited. These citations are easy to find by using Google Scholar, ResearchGate, Academic.edu, etc. Simply load these citations into a spreadsheet bar chart as I have illustrated in my personal history found in Figure 1.

This simple bar chart is a much more accurate portrayal of scholarly impact than the IF coefficient.

It includes all citations used in other articles, textbooks, scholarly monographs, policy manuals, dissertations, conference presentations, and any other manuscript available online. It could, in fact, include fraudulent material, but the Rank and Tenure Committee has all citations from the resume—unlike the IF coefficient. Such backup data have never been available for the IF coefficient. Pure quantification has its limits in assessing. Such a bar chart must include a qualitative¹ assessment. The inclusion of a qualitative analysis is not an available option with the IF coefficient.

After studying the construction and use of the IF, I have concluded that the IF coefficient is such a weak measure of an individual’s scholarly impact, it is unethical to employ it as a method of making rank and tenure decisions. In the study of research methods, we begin with an abstract concept. We take this abstraction and construct a measurement for it. A conventional abstraction we commonly measure is “intelligence.” We measure this abstraction by using the IQ test. Theoretically, intelligence comprises many components that are operationalized within subscales. Yet no IQ test includes all the theoretical subscales. Like most abstractions that we attempt



to measure in the social sciences, the IQ test fails to capture the totality of “intelligence” as a concept. Nevertheless, we use the IQ test a great deal in social science. In sociology and social work and in other academic circles, employment of the IQ is controversial. As questionable as the IQ test is, the use of IF coefficient for rank and tenure decision is *much worse*. The IF coefficient is not an ethical tool to use for personnel decision-making.

I see a lawsuit in the future. Psychometrists place measures like the IF coefficient in the category of a “gateway” instrument. By “gateway,” we mean the instrument influences the decision-making process on an individual’s livelihood. In contemporary psychometry, gateway instruments must include reliability and validity. Reliability is simple to assess. Greenwood (2007) assessed the reliability of the IF coefficient and found it weak. He writes, “the implications for advertisers, researchers, and journals is that only limited confidence can be placed on the ranking of these indicators. Decisions placed on such measures are potentially misleading...” (p. 52). For a measure to be valid, it must “measure what it purports to measure” and it “must not measure something other than what it purports to measure.” In his article defending IF, one of the creators, Garfield (2006), provides no coefficients that have or can statistically assess reliability or validity. *It is extraordinary and unprecedented* that a *gateway* instrument with a strong influence on the existence of one’s professional position fails to comply with statistical psychometric standards for reliability and validity. Failure to meet these standards can be easily demonstrated in court. I see a lawsuit in the future and hope that IF users and advocates have their malpractice insurance up to date. I am convinced that if a university’s legal counsel understood the mathematics embedded in the IF coefficient, the lawyer would realize that the university’s use of IF would be indefensible in court. In the unlikely event that such a complaint would appear on a court’s docket, punitive damages would be awarded to the faculty member who was denied promotion or tenure. That is my prediction.

I am interested in hearing from you regarding your opinion and use of IF coefficients in making rank and tenure decisions. I will publish your comments. Email me at smarson@nc.rr.com.

¹Unlike IF, use of the bar chart requires the faculty member to have access to all the citations that contribute to the bar chart. Rank and Tenure Committees have the power to assess these citations for a systematic qualitative assessment. This type of precise assessment is not available within the IF protocol.

References

- Garfield, E. (2006). The history and meaning of the journal impact factor. *JAMA: Journal of the American Medical Association*, 295(1), 90–93.
- Greenwood, D. C. (2007). Reliability of journal impact factor rankings. *BMC Medical Research Methodology*, 7, 48–53.
- Hoeffel, C. (1998). Journal impact factors. *Allergy*, 53(12), 1225.