COLLNET Journal of Scientometrics and Information Management
ISSN : 0973 – 7766 (Print) 2168 – 930X (Online)
Vol. 19(2) December 2025, pp. 171–195
DOI : 10.47974/CJSIM-2024-11003

A dozen topics to debate before using bibliometrics to evaluate scholarly research articles : A beginner's guide

M. Ryan Haley

Citation counts, and metrics built thereupon, have transformed how many colleges and universities evaluate the scholarly output of their faculty. Journal impact factors and scholar-level h-indices, among many other metrics, now appear in promotion, renewal, and tenure decisions as well as award and grant deliberations. Correctly using these seemingly objective measures of scholarly performance requires more tact and skill than many administrators and faculty realize. Moreover, the dizzying array of research evaluation metrics can befuddle even the most earnest of research evaluators. This paper concisely reviews twelve key considerations based on abridged theory and frontline bibliometric application. The intended audience is non-expert faculty and administrators that seek basic bibliometric guidance as they move into research evaluation roles.

JEL Classifications: *A1, M5.*

Keywords: *Scholar ranking, Journal ranking, Citations, Journal impact factor, h-index, g-index, e-index, Euclidean index, Citation database, Scientometrics, Informetrics, Citation cartels, Collusive citation, Research management, Google Scholar, Publish or perish, Web of science, Scopus.*

1. Introduction and Motivation

It is disconcertingly easy to misunderstand and accordingly mis-use bibliometrics indicators, which is cause for concern in the context of academic personnel decisions, workload allocation, grants, professorial awards, etc. (e.g., McKiernan et al., [1])[1]. These challenges compound

M. Ryan Haley
Department of Economics
University of Wisconsin
Oshkosh
800 Algoma Blvd.
Oshkosh
WI, 54901
U.S.A.

E-mail
haley@uwosh.edu

[1] While there are subtle differences between Bibliometrics, Scientometrics, and Informetrics, they are arguably synonyms [49] and so only the term Bibliometrics will be used hereafter for simplicity.

when research evaluation bodies have little or no formal bibliometric background, which is often; i.e., Bibliometrics is rarely (if ever) part of formal management training. Moreover, mastering the nuances and complexities of Bibliometrics requires a nontrivial degree of quantitative ability, which may not be in abundance on administrative research evaluation bodies.

While many bibliometric articles develop new methods, refine existing methods, or present field-specific empirical investigations; relatively fewer articles focus on frontline bibliometric guidance intended for non-expert faculty and administrators in academic settings. The latter is the focus herein, and therefore readers should imagine themselves as such a person (an "evaluator", hereafter), earnestly attempting to learn and apply basic Bibliometrics in an efficacious manner, free from deleterious incentives and unintended consequences.[2] In such a setting, where does the evaluator begin and what should the road map look like thereafter? This paper will hopefully help the evaluator orchestrate fruitful conversations with stakeholders, which will in turn answer these questions. To this end, the article is framed around twelve discussion topics the evaluator could (and probably should) explore when developing, deploying, directing, and updating a bibliometric evaluation process.[3]

To add context and specificity to the twelve discussion topics, insights from a first-hand implementation are referenced, which involved creating a large-scale journal list for a multi-disciplinary business college. These implementation insights and other key takeaways appear at the end of each section and subsection to expedite the reader's experience.

The remainder of the paper is organized into four sections. Section two discusses how to begin developing a research evaluation process. Section three contains the twelve primary points of discussion evaluators should consider. Section four outlines several additional considerations, among them a holistic, inclusive, and multi-dimensional approach to research evaluation that an evaluator may prefer over a single metric. Section five contains a summary and conclusions.

2. Where to Begin?

Where should the evaluation committee begin? A first question to ask is: Which specific types of scholarly output does the evaluator intend to evaluate? Is it books, monographs, technical reports, posters, showings, presentations, and/or research articles? The most

[2] Because of the intended audience, technical bibliometric details are withheld or relegated to footnotes. There are many good review sources, however, including Bornmann et al. [50], Wildgaard, Schneider, and Larsen [51], Waltman [52], and Aksnes et al. [7], to name a few.

[3] Similar efforts have appeared elsewhere in Bibliometrics; e.g., Bornmann et al. [53], Qiu et al. [54], Rousseau, Egghe, and Guns [55], Roldan-Valadez et al. [56], Ahmi [57], Donthu et al. [58].

common answer is research articles, which are principally measured using citation counts in some manner. Thus, citation-based evaluation of scholarly research articles will be the primary focus below.[4]

There are two primary article evaluation paradigms. The first evaluates the citation performance of a scholar's specific articles (hereafter, "micro evaluation"); the second evaluates the citation performance of the journals within which the scholar's articles appear (hereafter, "macro evaluation"). Which should the evaluator focus on? In many cases, this is a relatively easy decision because one paradigm or the other probably already philosophically dominates in the evaluator's field. For example, in fields like Economics and Business, there is a strong tradition of macro evaluation (e.g., [2]); in contrast, some science fields prefer micro evaluation. Hybrid approaches are also possible (e.g., [3]), which incorporate elements of micro and macro evaluation. If the evaluator is unclear which to choose, exploring the pros and cons of each would be an important first task (more on this in a later section). Regardless of the path chosen, a central focus for the evaluator will be some form of citation count (or function thereof); specifically, author-specific citation counts in micro evaluation and journal-specific citation counts (upon which journal impact factors are based) in macro evaluation.[5]

Key Takeaways and Implementation Insights

- Choose between author-level (micro) and journal-level (macro) assessment, or a hybrid thereof. Use field-specific practices to guide this choice.
- Understand that citations counts, whether they be at the author or journal level, will be the primary data needed to implement the evaluation process.
- Good citation data sources are a must; popular choices include Scopus, Web of Science, and Google Scholar. Search tools, such as Harzing's Publish or Perish software,[6] greatly facilitate the gathering of citation data and/or forming of metrics (more on this below).[7]

[4] The evaluation considerations set forth herein can be adapted to other forms of scholarly output, though each will have their own nuances. Also note that citation-based evaluation, a "revealed preference" approach, is not the only path forward; there is also the "stated preference" paradigm that relies on surveybased professional judgment of, say, a journal's rank as opposed to a journal's citation profile. See, for example, Bontis and Serenko [59].

[5] The bibliometric literature is thick with articles about the merits and shortcomings of micro vs. macro evaluation (using various naming conventions); see, for example, Seglen [60], [61], Oswald [62], Sgroi and Oswald [9], Bar-Ilan and Halevi [63], Larivi/ere and Sugimoto [64], Haley and McGee [65], and the DORA Declaration, to name a few of many. The goal herein is not to recount this debate, but to instead focus on helping evaluators ask the right questions and correspondingly induce productive discussions with stakeholders.

[6] Harzing [66] Publish or Perish, available from https://harzing.com/resources/publish-orperish

[7] Eigenfactor.org is a popular choice for macro evaluation; other sources also exist, which may be germane to a specific field or sphere of fields. For example, the Australian Business Deans Council (ABDC) journal list ranks business and economic journals.

3. Twelve Discussion Topics

The seemingly simple process of counting citations, whether at the micro or macro levels, is more complicated than it may appear. To help the evaluator think through the many challenges of their endeavor, assembled below are twelve discussion topics, each appearing in its own brief subsection. Several of the subsections include simple tabular demonstrations of the corresponding topic; these are meant to help frame the discussion, not solve the matter or recommend a specific solution. The tables typically presume the scholar attributes *not* mentioned are otherwise equal, to more cleanly demonstrate the subsection-specific topic. Each of the subsections is an incredibly brief introduction to the corresponding topic; expansive bibliometric literatures underlie each, which can be explored in more detail as the evaluator sees fit (some references are provided).

3.1 Knowledge, Patience, Inclusiveness, and Humility

One of the first challenges an evaluator will face, particularly in an administrative role, is the impulse to embrace quick research evaluation solutions. This "need for speed," which in turn fosters oversimplification, is a problematic characteristic that has not gone un-noticed by bibliometric researchers; e.g., Iglesias & Pecharromán [4] offer up an especially biting criticism of the *h*-index [5], a popular citation-based index:

In spite of all the previous caveats and qualifiers, it should be obvious that the *h*-index is going to be used to hand promotions and to allocate resources, despite [6] some research managers pious declarations that they purposely "avoid using impact factors and citation indices". And since the depths of politicians' (and administrators') "uniquely simple personalities" [10] only begins to be fathomed, it appears necessary to elaborate a bit on the portability of the *h* -index across the boundary between different scientific fields, lest it be grossly misused by eager specimens of the above sets.

This rather vivid critique may seem theatrical. However, the weight of the evaluator's responsibility becomes apparent when one considers the personnel, award, workload, and grant decisions that often hinge upon the evaluator's decisions. Because of this, it is often prudent and fruitful for the primary evaluator to work with a bibliometric development task force (or similarly charged body of stakeholders) to study the evaluation options, hold workshops, answer questions, etc. If need be, consultants can help get the process started and stay on track.

It is important to recognize at the outset that micro- and macro-level citations, despite their central role in Bibliometrics, are only able to imperfectly measure certain aspects of research [7]. Accordingly, it is common in bibliometric articles to humbly remind the reader that research evaluation should be multi-dimensional; i.e., that putting too much weight on a one-size-fits-all and/or "black box" research evaluation system will frequently result in nontrivial evaluation errors. Phelan [8] articulates this eloquently:

Bibliometric measures should be viewed as a useful supplement to other research evaluation measures rather than as a replacement.

Additionally, a good research evaluation process will typically be simple, but not simplistic.

Key Takeaways and Implementation Insights

- A metric (e.g., *h*-index) may not be comparable across fields without proper adjustments. Applying unmodified metrics across-the-board will often lead to grievous evaluation errors.
- Create a research evaluation process from an informed perspective, seeking multiple viewpoints and expert advice. Communicate the details of the process carefully.
- Sound evaluation processes should use a variety of metrics, including first-hand review of research accomplishments (when feasible), that weigh different aspects of scholarly output.
- Simplistic and/or hurriedly assembled evaluation processes will rarely work well.

3.2 *What is the Evaluation's Purpose?*

This is a crucial question to answer in the early stages of the evaluation development process. Best practices will depend on the evaluation goals. For example, evaluating articles from recent and/or short time frames (e.g., annual reviews, 2-3 year renewal contracts) will often be easier using macro evaluation, for the simple reason that scholar-specific citation counts take time to accumulate, whereas macro measures (like journal prestige) are immediately available at the time an article is accepted.[8] In contrast, a lifetime achievement award might sensibly weigh both micro and macro metrics; Sgroi and Oswald [9] make similar observations.

If the evaluator is struggling to clearly identify or articulate an evaluation's purpose, it may be useful to pose an explicit set of goals. For example, a goal might be to avoid incenting the overproduction of low-quality articles. Another goal might be valuing recent accomplishments because they reflect the current and near-future skills of the scholar. These are but two examples; "best" goals will vary by department and college and thus would be good topics to cover in the aforementioned workshops. Highly ranked departments and colleges, for instance, may only value articles published in elite journals when determining tenure, promotion, and professorial awards; if so, their evaluation process should logically reflect this goal.[9] In more modestly ranked departments and colleges with higher teaching demands, the list of acceptable publication outlets would be longer and accordingly require more nuanced goals.

[8] Altmetrics, which are discussed later, can potentially facilitate micro evaluations in constrained time frames.
[9] Heckman and Moktan [67] explore this in the field of Economics.

Key Takeaways and Implementation Insights

- Identify goals for the evaluation process. What does the evaluation body value? What type of scholarly output does it want to incent?
- Recognize that macro assessments may be more useful for short-term evaluations.
- Evaluations with long time horizons, (e.g., decades) may prefer micro (or hybrid) assessments because the author-level citation counts have had time to accumulate.

3.3 *Inter-Field Citation Comparisons*

Can citation counts be compared in an "apples-to-apples" manner for scholars in different fields, or are inter-field adjustments necessary? The quote from Iglesias & Pecharromán [4] above points to this issue, which has been studied at length in Bibliometrics.[10] This debate has several fronts:

Table 1

Scholar Comparison Example: Inter-Field Issues

Field	Max Field *h*-index	Scholar	Scholar-Specific *h*-index
A	90	1	30
B	60	2	30
C	30	3	30

- Some fields and sub-fields are inherently smaller than others and accordingly have fewer scholars and journals and therefore inherently lower citation potentials, which correspondingly impacts micro and macro evaluations. Consider the example in Table 1, which contains scholars (1-3) from each of three distinct fields (A-C). The maximum *h*-index for each field is reported as well.[11] The scholar-specific *h*-indices (all equal to 30) suggest the scholars are equally accomplished. However, when the scholar-specific *h*-indices are compared to the maximum *h*-index in their respective fields, Scholar 3 appears to be at the very top of their field whereas Scholars 1 and 2 do not. Evaluators may seek ways to recognize and potentially adjust for this bibliometric reality. [12]

[10] See, for example, Pudovkin and Garfield [68] Iglesias and Pecharrom/an [4], Radicchi, Fortunato, and Castellano [69], Moed [70], Leydesdorff, Zhou, and Bornmann [71], Waltman and van Eck [72], Ioannidis et al., [4], Haley and McGee [65], or Haley [17].

[11] The h-index [8] is among the most widely known bibliometrics used in micro and macro evaluations. It is probably easiest to understand the h-index using simple numeric examples. If a scholar has an h-index of 10, this means their 10 most highly-cited articles each have at least 10 citations. Similarly, if a journal has an h-index of 300, this means its 300 most highly-cited articles each have at least 300 citations.

[12] At times, scholars may narrowly define their field to appear exceptional. To avoid this type of gaming, the evaluator may want to discuss how to best govern the definition of a scholar's field.

- A corollary to the prior item concerns whether there may be cause to distinguish between a scholar that is a "big fish" in a small/shallow pond (e.g., a popular but simplistic research area) vs. a scholar that is an "average fish" in a big/deep pond (e.g., a complex field with less visibility)? Referring again to Table 1, it may be that Field A (with the largest field-specific *h*-index) is a simplistic field whereas Field B is especially complex and difficult to publish in. Discussing the simplicity vs. complexity of a field or a scholar's research dossier is often difficult because scholars tend to think of their own field and work as superior. Traditional peer-review evaluation processes may be better able to make these types of distinctions than "objective" metrics (e.g., *h*-index). This possibility, taken alongside the prior note above, evinces how there can be an array of subtle and potentially contradictory aspects in the evaluation process.[13]
- Asymmetries in average citation counts and/or metric scores can result from citation culture differentials or database attributes; i.e., in some fields it is typical to have exhaustive bibliographies whereas other fields less so (e.g., Leydesdorff, [11]; Crespo, Li, and Ruiz–Castillo [12]). Why such differentials exist has also been studied; e.g., Marx and Bornmann [13]. Regardless of their genesis, inter-field citation differentials can lead to the cases discussed in the prior points, wherein a field has inflated scores because of its citation culture, not necessarily because the research is higher quality or more impactful. Put another way, if a field perceives its relative value as a direct function of its citations, then it has an incentive to pursue low-cost ways to inflate citation counts. In many cases there is little to restrict such a practice, other than a field's publishing decorum. This further underscores the importance of discussing inter-field comparison issues.

Key Takeaways and Implementation Insights

- Citation counts and citation metrics (e.g., *h*-index) are difficult to compare across fields for a variety of reasons, some of which are outlined above.
- Without field-specific adjustments, citations counts and metrics built thereupon can lead to very flawed research comparisons across scholars/fields, both at the micro and macro assessment levels.
- Different fields often have different citation practices, which in turn can influence the evaluation process.
- Fields often differ in size, as measured by scholar and journal count.
- Some fields prefer fewer large-scale works where as others prefer more frequent smaller-scale contributions.

[13] It could also be argued that simpler fields, howsoever defined, may have an inherent citation advantage because a larger percentage of scholars can participate in these fields, whereas very complex fields are only accessible to the upper-most tail of the scholar distribution.

- Database choice can favor one field over another by dint of their coverage.

3.4 *Older vs. Newer Articles*

One irrefutable fact about citation counts is that older articles have had more time to accrue citations than newer articles. Moreover, the rate at which citations accrue can vary greatly across fields, sub-fields, and even at the article-specific level. Thus, if the evaluator plans to compare the citation counts of a 5-year scholar vs. those of a 20-year scholar, the latter will have a significant advantage, all else constant. To demonstrate, consider Table 2, which contains citation data for two scholars with identical citation and article counts. Based on raw citation count alone, these two scholars appear to be equally accomplished (i.e., both have 100 total citations), but one may wonder if such a conclusion is rationally just for Scholar 1. Dividing article-specific citations by the article's age results in a per-year citation count per article.[14] Aggregating these ratios makes Scholar 1 appear more accomplished than Scholar 2 (i.e., 43 vs. 13). Article age adjustment is most germane when evaluating scholars over differing frames.

Table 2

Scholar Comparison Example: Article Age Adjustment

Article	Scholar 1:			Scholar 2:		
Rank	Citations	Age	Citations/Year	Citations	Age	Citations/Year
1	40	5	8	40	10	4
2	30	2	15	30	6	5
3	20	2	10	20	10	2
4	10	1	10	10	5	2
Totals	100		43	100		13

Key Takeaways and Implementation Insights

- It is often necessary to adjust for article age when comparing scholars with different tenures.
- Differing scholar ages can also affect popular author-specific citation-based indices like the *h*-index, *g*-index [14], and *e*-index [15], among others.[15]

3.5 *Sole vs. Multi-Author Articles*

How do citations of multi-author papers compare to citations accumulated by sole-authored papers? To see this issue more clearly, consider Table 3. Note that both scholars have the same number of articles and citations, but differ on the number of authors per

[14] This is perhaps the most obvious way to perform article age adjustment [17], though other approaches exist.
[15] Merged metrics also appear in Bibliometrics; e.g., Alonso et al. [73].

paper.[16] Further assume all else is equal such as scholar age, journal quality, etc. Dividing the citation count per article by the number of authors [16], [17] is one possible way to rationally adjust the raw citations, though other adjustment processes exist (e.g., Perianes-Rodriguez, Waltman, and Van Eck [18]). Put another way, does the evaluator believe that assigning full citations to each author is a logical representation of the workload needed to complete the article?

Table 3

Scholar Comparison Example: Sole vs. Multiple Authors

Article	Scholar 1:			Scholar 2:		
Rank	Citations	Authors	Citations/Author	Citations	Authors	Citations/Author
1	40	1	40	40	4	10
2	30	1	30	30	5	6
3	20	1	20	20	2	10
4	10	2	5	10	1	10
Totals	100		95	100		36

Not adjusting for author count can induce some peculiar incentives. As a stark example, suppose a department has ten scholars, each of whom writes a single paper and then adds the other nine scholars as authors. Now, each scholar appears to have written ten articles, when in fact they have each written only one. Moreover, without fractional citation counting of some type, each scholar will accumulate citations on nine articles to which they made negligible contributions. This is but one of the many types of metric gaming that can arise if the evaluator does not consider a metric's incentives.

A related issue is the ordering of author names, which has an array of meanings depending on the field (e.g., Ioannidis, Boyack, and Wouters, [19]). In some fields, alphabetical order is the standard; in other fields, the author order implies the rank level of contribution; in other fields, the order indicates seniority or type of contribution, etc. This reality is a) something the evaluator needs to be aware of and b) something that may need to be formally accounted for before clean evaluations are possible. Author ordering differentials can sometimes surprise neophyte evaluators who may implicitly assume that all fields assign authorship in the way with which they are familiar.

Key Takeaways and Implementation Insights

- It is often necessary to adjust for sole vs. multi-athorship to create sound workload-based scholar comparisons.
- Undesirable incentives can arise in the absence of fractional citation counting.
- Different fields have different author-ordering conventions, which often warrants consideration before making comparisons across authors.

[16] In some fields it is not uncommon for articles to have a dozen or more authors.

3.6 *Citation Differentials Across Article Type*

Certain types of research articles accumulate higher citations counts, on average, than other types. Theoretical articles tend to be cited the least, applied papers more, and review articles the most [20], [21]. The degree of differentiation varies by field, but in some research areas review articles are cited nearly seven times more often than regular articles [22]. Thus, comparing a theorist's citation count to a review article aficionado may not be an "apples-to-apples" comparison. Additionally, if the evaluator values only raw citations, that incents scholars to write review articles instead of original research, which may or may not meet a university's research goals.

Four related issues arise in the discussion of article type. First, publishing articles as open access vs. behind a paywall can systematically influence citation rates; relatedly, scholars with more professional funding are better able to cover costs of open-access publishing. Second, the structure and length of an article's title and similar superficial factors can impact citation performance [23], [24], [25]. Third, average article length (e.g, full length articles vs. letters) can vary by field, sub-field, and scholar, and therefore may be cause for evaluator attention, particularly when article count is central to the evaluation process. Fourth, journals are predominately written in English, which creates challenges for scholars in non-English speaking countries (e.g., [26]).

Key Takeaways and Implementation Insights

- Different types of articles, such as those discussed above, tend to have significantly different average citation rates. Adjusting for this in some way may be warranted.
- Adjustments for funding, language, and article length may be necessary to create fair comparisons between scholars.

3.7 *Depth vs. Breadth Relevance*

Perry and Reny [27] argue that "depth relevance" is an important premise upon which a satisfactory evaluation metric should be built. The idea is to value a smaller number of highly cited papers over a larger number of more sparsely cited papers. Depth relevance rewards "one-hit-wonders" in a way that may not be perceived as rationally just to scholars that regularly publish articles that earn average citation counts. Depth relevance also implicitly assumes that lengthy, thorough articles are a more effective way to transmit knowledge than a larger number of shorter letter-style articles; whether this is true probably varies by field.

Breadth relevance is a competing concept (e.g., [28]), which argues that a research dossier with more balance in its article-specific citation count should be preferred.[17] However, it is possible to create generalized metrics that can flex into depth or breadth

[17] It is perhaps worth noting that total raw citation counts lies exactly between depth and breadth relevance; i.e., it satisfies neither premise.

relevance to whatever degree the evaluator desires; see Stern and Tol [28] for more details. The evaluation goals should reflect which type of relevance the evaluator values.

Breadth relevance has a second possible meaning in Bibliometrics, which revolves around the question: How may truly distinct articles has a scholar written? Some scholars may recycle a data set for numerous articles that are only slightly different (e.g., [29]); similarly, some scholars may use the same method and just change out the data section over and over across articles. In contrast, other scholars may write truly distinct articles that have minimal or no overlap with prior data or methods they have used; i.e., some scholars may aspire to be more a polymath than a repetitive specialist. Polymath aspirations constitute an especially ambitious scholarly endeavor because they entail frequently learning new literatures, new methods, new data, and new editorial expectations. Both type of scholars have cause to perceive their accomplishments as valuable; the research evaluation goals should articulate which the evaluator prefers (or if the evaluator is indifferent).

Key Takeaways and Implementation Insights

- Some fields may prefer depth relevance whereas others may prefer breadth relevance. Understanding each and which better fits the evaluator's vision and goals is important during the policy build.
- Interdisciplinary departments may be especially interested in breadth relevance.
- Metrics exist that allow evaluators to adjust the degree of depth vs. breadth relevance.

3.8 *Citation Databases, Citation Purity, and Self-Citations*

There are array of different citation databases from which the evaluator can choose. Examples include Google Scholar, Scopus, Web of Science, to name a few. Citation databases are not necessarily created equal; indeed, there is an extensive bibliometric literature on the pros and cons of different citation databases, how their coverage can vary by field, and how authors may try to game them (e.g., [6], [30], [31], [32], [33], [34], [35]).

A related concern is the cleanliness of the citation data. Have the citations been vetted for errors of omission and commission? For example, searching for scholar-specific citations (in a micro evaluation) using a common surname will frequently return polluted citation data and will, accordingly, pollute citation metric scores. Some authors may have changed surnames over the course of their careers; if so, how should that be handled? Journals, even major journals, will occasionally change names as well. Macro evaluation enjoys a law of large numbers-like effect in that errors of omission and commission are more likely to average out over the many different articles and scholars in a journal. Additionally, a journal is less prone to scoring errors when it is queried using its International Standard Serial Number (ISSN) as opposed to its title.

A common discussion in micro and macro evaluation concerns the handling of self citations [36], [37]. Should they be included in the evaluation process or not? Some citation

sources are careful to remove self citations (e.g., Eigenfactor.org) whereas other sources do not. The evaluator will need to decide which avenue to pursue, and if a satisfactory data source matches their needs. It should be noted that some scholars take self-citation to an extreme [38], and that self-citation increases the rate of non-self-citations [39]. This suggests that a scholar may be able to self-induce a Matthew effect by engaging in extensive self-citation.[18]

Key Takeaways and Implementation Insights

- Different citation data bases exist, which do not necessarily offer equal coverage across all fields.
- Citation data is often polluted, especially for scholars with common surnames. Failing to clean the citation data before use can dramatically impact scholar index values.
- Citation databases may have idiosyncratic errors of omission or commission, either of which can be challenging to fix.
- Journals change names, as do scholars at times. Adjusting for both may be necessary.
- ISSN is usually a cleaner way to query a journal than using the journal's title.
- Evaluators will have to decide how they want to handle self-citations.

3.9 *Coercive Citation and Citation Cartels*

Coercive citation occurs when a journal editor or senior colleague uses their position of power to pressure authors into citing specific articles [40, 41, 42]. This can artificially inflate citation scores at the micro and macro levels (e.g., [43], [44]). Some journals have taken a public stand against this behavior, though coercive citation remains a concerning trend.[19]

Citation cartels are informal agreements among scholars in a social network wherein member scholars monitor each others' citation metrics and strategically cite each others' research to maximize the growth rate of members' citation metric scores. Unlike a traditional economic cartel, there is little incentive for citation cartel members to defect [42] and they can be difficult to detect, which together suggests they can persist for extended periods.

Key Takeaways and Implementation Insights

- Evaluators should be aware of possible coercive citation practices within their institution.

[18] In Bibliometrics, the Matthew effect describes how a popular scholar, article, or journal tends to get disproportionately more citations [74], holding fixed the quality of the article. In plainer terms, the Matthew effect revolves around the concept of cumulative advantage; i.e., advantage tends to induce additional advantage at an increasing rate [75], [76]. The Matthew effect can also artificially inflate the citation count of "one-hit-wonder" scholars and tends to disproportionately benefit scholars that start their careers earlier [77]

[19] http://rfssfs.org/files/2012/09/EditorsJointPolicy.pdf

- Evaluators may consider disallowing journals than have a reputation for practicing coercive citation.
- Citation cartels can be difficult to detect. Using metrics that are less susceptible to citation cartels is an option.

3.10 *Predatory Journals*

It is important to recognize that some journals do not have meaningful review processes, and will frequently accept articles without formal or rigorous review, even if their stated review process indicates otherwise. Grudniewicz et al. [45] offer the following concise definition:

Predatory journals and publishers are entities that prioritize self-interest at the expense of scholarship and are characterized by false or misleading information, deviation from best editorial and publication practices, a lack of transparency, and/or the use of aggressive and indiscriminate solicitation practices.

It is perhaps worth adding that predatory journal aggressive solicitation practices often include the promise of very quick "review" time lines. Predatory journals have proliferated in the last decade, and now some have amassed relatively high impact factors and *h*-index scores. Thus, it does take some thoughtful diligence to keep them out of the evaluation process. There are several resources now, such as Cabells, Beall's list, and the Australian Dean List [20] (among others), that can help evaluators detect and exclude predatory journals; Grudniewicz et al. [45] concisely explores some of these options.

The rise of predatory journals creates some pressure for institutions to more carefully scrutinize the journals within which their scholars publish in. Journals with long and respected tenures are logical places to find "high ground" in the contemporary flood of predatory journals.

Key Takeaways and Implementation Insights

- Predatory journals often have low editorial standards and may have hefty Article Processing Charges (APCs) to accumulate profits.
- Some predatory journals may have craftily amassed significant impact factors, which can make them seem like reputable journals.
- Cabells and Beall's list are good ways to help detect predatory journals.
- It is important for evaluators to detect these types of articles/journals and bar them from the research assessment process.

3.11 *Workarounds*

A single metric or narrowly defined evaluation process will rarely offer rational justice to all scholars. It is therefore important to have codicils that permit auxiliary evaluation

[20] https://abdc.edu.au/research/abdc-journal-quality-list/

steps when warranted. For example, it may be prudent to establish a workaround for adding a new journal to a journal list. As a second example, certain scholars may have quantitative or qualitative elements to their research that mainstream citation metrics do not measure fully or at all; e.g., news attention to a published article, working paper downloads, conferences, or the positive institutional effects precipitating from a scholar's white paper or technical report.[21]

When debating a workaround policy, the evaluator will quickly realize that scholars have a tendency to argue in favor of their personal accomplishments and pre-existing bibliometric habits. For example, a scholar that uses self-citations extensively will advocate for their inclusion in the evaluation process. A scholar that publishes in reputed journals will often campaign for macro evaluation (or hybrid). Scholars who have not published in reputed journals will argue that journal reputation should not matter. A sound workaround system can serve as an effective catch basin for these types of concerns and complaints, assuming the evaluator sees merit in making any accommodations.

Key Takeaways and Implementation Insights

- Research evaluation processes will need formal workarounds for conditions that cannot be adequately captured by citation counts, author metrics, and/or journal metrics.
- Scholars will tend to favor their prior behaviors when developing a research evaluation process.
- Evaluators should build a process that directly incents the specific research goals of the college/university.

3.12 *Incentives and Phasing*

Any specific evaluation process will spawn a network of incentives. Some will be intentional, easy to see, and positive for scholars and the institution; others will be unintended, potentially difficult to see ex ante, and may well incent behavior that is inconsistent with the larger goals of the faculty or administration. Thus, it is prudent to create a deliberate and thorough "incentive map" wherein the evaluator tries to anticipate the array of behavioral responses a proposed evaluation process will induce. A sound understanding of the proposed evaluation metrics and any concomitant workarounds will expedite and facilitate this process. A useful rhetorical device when mapping incentives is to imagine an especially difficult, passive-aggressive scholar that seems to delight in skirting rules; viewing a proposed evaluation process through this lens may seem unsavory, but it can help insulate the process from the many forms of citation manipulation and gaming [46].

[21] A simple bibliometric example might be an article that establishes an authoritative journal list for a specific field [33]; such an article may get viewed, downloaded, and presumably used extensively by faculty and evaluators in that field, but may not be cited extensively in academic articles. In a similar vein, articles that are perceived by scholars to be of high quality may not be perceived as such by front-line practitioners [78]

Any newly minted evaluation system will create ex post winners and loser, which may warrant "grandfathering" to avoid disenfranchising scholars who were researching under a different set of incentives prior to the new policy. Accordingly, it is logical to debate the pros vs. cons of incrementally phasing in the new paradigm so scholars have time to adjust their research processes to align with any new incentives.

Key Takeaways and Implementation Insights

- All metrics create various types of incentives. Tracing them through as best as possible during the building of the research evaluation process will help engender success.
- Be prepared to update a process 1-3 years after implementation; rarely will it be perfect from the start.
- Imagine a disagreable faculty member to help trace out how a research evaluation system may be prone to gaming or manipulation.
- Grandfathering and phasing in a new process may greatly increase the buy-in and the success of a new research evaluation process.

4. Additional Considerations

The section covers several topics the evaluator may want to consider, including some guidance on how to choose between micro vs. macro evaluation (or perhaps choose both), a brief discussion of Altmetrics, and a brief presentation of a flexible holistic evaluation option that may be attractive to some evaluators that value a wide array of scholarly accomplishments.

4.1 *Choosing Between Micro vs. Macro Evaluation*

In some cases an evaluator will need to decide whether to pursue micro or macro evaluation. Several litmus-like tests are proposed below to help the evaluator (and evaluation committee, more generally) recognize their preferences and discuss from there.

First, consider Table 4, which contains citation information for two scholars with identical citation counts (and all else constant). The only difference is that Scholar 1's article were published in elite-level journals where Scholar 2's articles appeared in more modestly ranked journals. Based on raw citation count, the scholars appear to be equally accomplished (i.e., both have 100 total citations). If the evaluator agrees with this, they will probably favor micro evaluation. If the evaluator sees Scholar 1's accomplishments as more impressive, they will likely favor macro evaluation.[22]

[22] Again, hybrid approaches are possible; see, for example, Haley [17], [79], [80] Additionally, Sgroi and Oswald [9] discuss when micro vs. macro evaluation may be preferred.

Table 4

Scholar Comparison Example: Journal Prestige

Article	Scholar 1:		Scholar 2:	
Rank	Citations	Journals	Citations	Journals
1	40	Elite Journal A	40	Modest Journal W
2	30	Elite Journal B	30	Modest Journal X
3	20	Elite Journal C	20	Modest Journal Y
4	10	Elite Journal D	10	Modest Journal Z
Totals	100		100	

A second litmus-like test might discuss recent experimental evidence that highlights how adding low-level publications to an otherwise high-level publication list actually lowers overall ratings of the scholar [47]. If the evaluator finds these results compelling, they may favor macro evaluation.

A third litmus-like test centers on the nature of the evaluation process. Is the primary purpose to evaluate recent scholar accomplishments? The "what-have-you-done-for-me-lately" evaluation focus is common, matching to the practical realities of annual reviews, biennial merit scoring, 2-3 year contract renewal intervals, and post-tenure review windows. In most of these cases, micro evaluation will often be under-developed for the most recent scholarly research accomplishments, which may make macro evaluation a more workable focus.

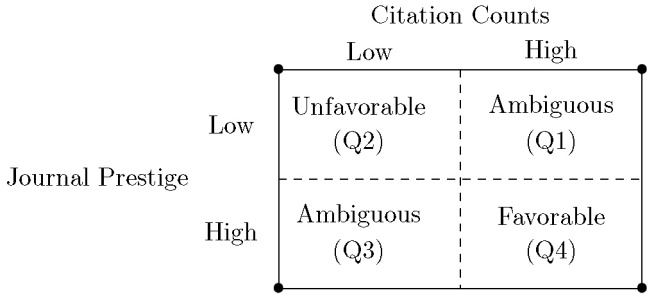

Figure 1

Four Scholarly Classifications

A fourth litmus-like test can be conceptualized around Figure 1, which contains a simple binarized schematic for categorizing a scholarly article (or scholar, more generally).[23] Most evaluators would probably be comfortable rating articles with low citation counts in

[23] The schematic is binarized for simplicity, more granular divisions are certainly possible. Note that the individual evaluator would need to tangibly define what is meant by "Low" vs. "High" for Journal Prestige and Citation Counts.

low-level journals (Q2) less favorably than articles with high citation counts in high-level journals (Q4). How to value articles that fall into Q1 and Q3 is probably less clear. If an evaluator sees minimal value in Q3 articles, they will probably prefer micro evaluatement; if instead they see little value in Q1 articles, they will probably prefer macro evaluatement; and if they see a possible equivalence or sliding trade-off between Q1 and Q3, then they may prefer a hybrid evaluation approach that accommodates both micro and macro successes.

Key Takeaways and Implementation Insights

- Choosing between micro- and macro-level research assessment can often be accomplished by referencing established field-specific practices.
- Some advice on how to choose between micro vs. macro assessment paradigms is provided for cases where field-specific practices are not workable.
- Macro assessment can be especially useful for short term, time-constrained evaluations because the prestige of a journal is fully known at the time of article acceptance whereas how many citations the article will ultimately achieve lies years in the future.
- Macro assessment may be useful as a signal of research ability and effort in fields where prestigious journals are perceived to be much more difficult to publish in due to low acceptance rates, referee rigor, and high contribution standards.

4.2 *Alternative Outputs and Altmetrics*

The discussions above have primarily centered on article evaluation, but other forms of measurable scholarly output of course exist (e.g., books, working papers, technical reports, conference proceedings). More still, citations and citation-based metrics also have competitors, among them Altmetrics, which are alternative ways to measure the attention and influence of scholarly output. Examples include abstract views, downloads, Mendeley bookmarks, news coverage, blog coverage, and tweets. Altmetrics can nicely augment micro evaluations because they typically develop more quickly than citations[24].[24]

Key Takeaways and Implementation Insights

- The focus herein has been on peer-reviewed journal articles, but other types of scholarly output may be reasonably included in a research evaluation process.
- Altmetrics can be useful and timely ways to augment research assessment (especially micro assessment).

5. Summary and Conclusions

The purpose of this paper was to address the neophyte research evaluator with minimal bibliometric knowledge, but who has nonetheless been charged with creating a scholar

[24] For a informative discussion of Altmetrics, see, for example, Piwowar [81] Bornmann [82] or Didegah, Bowman, and Holmberg [83].

evaluation process for their department or college. The narrative was deliberately void of technical details, complicated schematics, and personal views of the author; all with the hope of making the article more accessible to and informative for the non-expert reader. Bibliometrics can seem simple at first glance, but like most fields, complexities and nuances arise upon closer inspection. The twelve discussion points set forth herein will hopefully help focus some of the evaluator's conversations with stakeholders. The evaluator that begins from an informed perspective will be miles ahead of an evaluator who simply tries to intuit or fast-track their way through the process. Ignoring the issues outlined above will almost surely result in a fretful and contentious evaluation process.

The new-to-Bibliometrics reader may be overwhelmed by the sheer volume of evaluation considerations discussed above. Indeed, each topic mentioned briefly herein is but the very tip of its own expansive literature. However, there is hope. Article evaluation metrics can be managed [48] but doing so will almost surely entail a greater effort level than most evaluators expect. However, when one considers all the far-reaching implications of the research evaluation process (e.g., tenure, contract renewal, awards, workload, meeting accreditation standards), creating and maintaining a sound evaluation process is of central importance to the vitality of academic institutions. This article does not outline every conceivable bibliometric discussion that should be considered, just the most common. Each application will have its own field-specific, department-specific, and/or scholar-specific nuances that will warrant deliberation. Bibliometrics perenially warns that the final evaluation of a scholar should not lean too heavily on a single metric; research evaluation is an inherently multi-dimensional endeavor that requires holistic consideration of the issues outlined in this article. This is arguably the most important lesson for the evaluator that aspires to ethical and responsible assessment: simplistic and/or mono-assessment practices will probably do more harm than good.

Conflict of Interest Statement

The author asserts that there are no interests or relationships, financial or otherwise, that might be perceived as influencing the author's objectivity in the writing of this article.

References

[1] E. C. McKiernan, L. A. Schimanski, C. Muñoz Nieves, L. Matthias, M. T. Niles, and J. P. Alperin, "Use of the Journal Impact Factor in academic review, promotion, and tenure evaluations," *Elife*, vol. 8, e47338 (2019). [Online]. Available: https://doi.org/10.7554/eLife.47338

[2] A. Attema, W. Brouwer, and J. Van Exel, "Your right arm for a publication in AER?" *Economic Inquiry*, vol. 52, no. 1, pp. 495–502 (2014). [Online]. Available: https://doi.org/10.1111/ecin.12013

[3] M. Haley and M. McGee, "Jointly valuing journal visibility and author citation count: An axiomatic approach," *Journal of Informetrics,* vol. 14, no. 1, art. 100988 (2020). [Online]. Available: https://doi.org/10.1016/j.joi.2019.100988

[4] J. Iglesias and C. Pecharromán, "Scaling the h-index for different scientific ISI fields," *Scientometrics,* vol. 73, pp. 303–320 (2007). [Online]. Available: https://doi.org/10.1007/s11192-007-1805-x

[5] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proceedings of the National Academy of Sciences,* vol. 102, no. 46, pp. 16569–16572 (2005). [Online]. Available: https://doi.org/10.1073/pnas.0507655102

[6] R. Ball and D. Tunger, "Science indicators revisited – Science Citation Index versus SCOPUS: A bibliometric comparison of both citation databases," *Information Services & Use*, vol. 26, no. 4, pp. 293–301 (2006). [Online]. Available: https://doi.org/10.3233/ISU-2006-26404

[7] D. W. Aksnes, L. Langfeldt, and P. Wouters, "Citations, citation indicators, and research quality: An overview of basic concepts and theories," *Sage Open*, vol. 9, no. 1, pp. 1–11 (2019). [Online]. Available: https://doi.org/10.1177/215824401982957

[8] T. J. Phelan, "A compendium of issues for citation analysis," *Scientometrics,* vol. 45, no. 1, pp. 117–136 (1999). [Online]. Available: https://doi.org/10.1007/BF02458472

[9] D. Sgroi and A. Oswald, "How should peer-review panels behave?" *The Economic Journal,* vol. 123, no. 570, pp. F255–F278 (2013). [Online]. Available: https://doi.org/10.1111/ecoj.12070

[10] V. Anauati, S. Galiani, and R. Gálvez, "Quantifying the life cycle of scholarly articles across fields of economic research," *Economic Inquiry*, vol. 54, no. 2, pp. 1339–1355 (2016). [Online]. Available: https://doi.org/10.1111/ecin.12292

[11] L. Leydesdorff, "Caveats for the use of citation indicators in research and journal evaluations," *Journal of the American Society for Information Science and Technology,* vol. 59, no. 2, pp. 278–287 (2008). [Online]. Available: https://doi.org/10.1002/asi.20743

[12] J. A. Crespo, Y. Li, and J. Ruiz–Castillo, "The measurement of the effect on citation inequality of differences in citation practices across scientific fields," *PLoS One*, vol. 8, no. 3, e58727 (2013). [Online]. Available: https://doi.org/10.1371/journal.pone.0058727

[13] W. Marx and L. Bornmann, "On the causes of subject-specific citation rates in Web of Science," *Scientometrics*, vol. 102, no. 2, pp. 1823–1827 (2015). [Online]. Available: https://doi.org/10.1007/s11192-014-1499-9

[14] L. Egghe, "Theory and practise of the g-index," *Scientometrics,* vol. 69, no. 1, pp. 131–152 (2006). [Online]. Available: https://doi.org/10.1007/s11192-006-0144-7

[15] C. T. Zhang, "The e-index, complementing the h-index for excess citations," *PLoS One,* vol. 4, no. 5, p. e5429 (2009). [Online]. Available: https://doi.org/10.1371/journal.pone.0005429

[16] A. Belikov and V. Belikov, "A citation-based, author- and age-normalized, logarithmic index for evaluation of individual researchers independently of publication counts," F1000Research, vol. 4 (2015). [Online]. Available: https://doi.org/10.12688/f1000research.6033.2

[17] M. Haley, "An EigenFactor-weighted power mean generalization of the Euclidean Index," *PLoS ONE,* vol. 14, no. 2, e0212760 (2019a). [Online]. Available: https://doi.org/10.1371/journal.pone.0212760

[18] A. Perianes-Rodriguez, L. Waltman, and N. J. Van Eck, "Constructing bibliometric networks: A comparison between full and fractional counting," *Journal of Informetrics*, vol. 10, no. 4, pp. 1178–1195 (2016). [Online]. Available: https://doi.org/10.1016/j.joi.2016.10.006

[19] J. P. A. Ioannidis, K. Boyack, and P. F. Wouters, "Citation metrics: A primer on how (not) to normalize," *PLoS Biology*, vol. 14, no. 9, e1002542 (2016). [Online]. Available: https://doi.org/10.1371/journal.pbio.1002542

[20] D. W. Johnston, M. Piatti, and B. Torgler, "Citation success over time: Theory or empirics?" *Scientometrics,* vol. 95, no. 3, pp. 1023–1029 (2013). [Online]. Available: https://doi.org/10.1007/s11192-012-0910-7

[21] J. Angrist, P. Azoulay, G. Ellison, R. Hill, and S. F. Lu, "Inside job or deep impact? Extramural citations and the influence of economic scholarship," *Journal of Economic Literature*, vol. 58, no. 1, pp. 3–52 (2020). [Online]. Available: https://doi.org/10.1257/jel.20181508

[22] R. Miranda and E. Garcia-Carpintero, "Overcitation and overrepresentation of review papers in the most cited papers," *Journal of Informetrics*, vol. 12, no. 4, pp. 1015–1030 (2018). [Online]. Available: https://doi.org/10.1016/j.joi.2018.08.006

[23] H. R. Jamali and M. Nikzad, "Article title type and its relation with the number of downloads and citations," *Scientometrics*, vol. 88, no. 2, pp. 653–661 (2011). [Online]. Available: https://doi.org/10.1007/s11192-011-0412-z

[24] M. Van Wesel, S. Wyatt, and J. ten Haaf, "What a difference a colon makes: How superficial factors influence subsequent citation," *Scientometrics*, vol. 98, no. 3, pp. 1601–1615 (2014). [Online]. Available: https://doi.org/10.1007/s11192-013-1154-x

[25] B. Deng, "Papers with shorter titles get more citations," *Nature*, vol. 2, no. 8, article 150266 (2015). [Online]. Available: https://doi.org/10.1038/nature.2015.18246

[26] V. Ramírez-Castañeda, "Disadvantages in preparing and publishing scientific papers caused by the dominance of the English language in science: The case of Colombian researchers in biological sciences," *PLoS ONE,* vol. 15, no. 9, e0238372 (2020). [Online]. Available: https://doi.org/10.1371/journal.pone.0238372

[27] M. Perry and P. Reny, "How to count citations if you must," *The American Economic Review,* vol. 106, no. 9, pp. 2722–2741 (2016). [Online]. Available: https://doi.org/10.1257/aer.20140850

[28] D. I. Stern and R. S. Tol, "Depth and breadth relevance in citation metrics," *Economic Inquiry,* vol. 59, no. 3, pp. 961–977 (2021). [Online]. Available: https://doi.org/10.1111/ecin.12994

[29] M. Errami and H. Garner, "A tale of two citations," *Nature,* vol. 451, no. 7177, pp. 397–399 (2008). [Online]. Available: https://doi.org/10.1038/451397a

[30] J. Beel and B. Gipp, "Academic search engine optimization (ASEO): Optimizing scholarly literature for Google Scholar and Co.," *Journal of Scholarly Publishing*, vol. 41, no. 2, pp. 176–190 (2010). [Online]. Available: https://doi.org/10.3138/jsp.41.2.176

[31] A. Diem and S. C. Wolter, "The use of bibliometrics to measure research performance in education sciences," *Research in Higher Education*, vol. 54, no. 1, pp. 86–114 (2013). [Online]. Available: https://doi.org/10.1007/s11162-012-9264-5

[32] M. Haley, "Ranking top economics and finance journals using Microsoft Academic Search versus Google Scholar: How does the new Publish or Perish option compare?" *Journal of the American Society for Information Science and Technology,* vol. 65, no. 5, pp. 1079–1084 (2014). [Online]. Available: https://doi.org/10.1002/asi.23080

[33] M. Haley, "A ranking of journals for the aspiring health economist," *Applied Economics*, vol. 48, no. 18, pp. 1710–1718 (2016). [Online]. Available: https://doi.org/10.1080/00036846.2015.1105927

[34] A. W. Harzing, "Two new kids on the block: How do Crossref and Dimensions compare with Google Scholar, Microsoft Academic, Scopus and the Web of Science?" *Scientometrics,* vol. 120, no. 1, pp. 341–349 (2019). [Online]. Available: https://doi.org/10.1007/s11192-019-03114-y

[35] J. Baas, M. Schotten, A. Plume, G. Côté, and R. Karimi, "Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies," *Quantitative Science Studies,* vol. 1, no. 1, pp. 377–386 (2020). [Online]. Available: https://doi.org/10.1162/qss_a_00019

[36] M. Seeber, M. Cattaneo, M. Meoli, and P. Malighetti, "Self-citation as strategic response to the use of metrics for career decisions," *Research Policy,* vol. 48, no. 2, pp. 478–491 (2019). [Online]. Available: https://doi.org/10.1016/j.respol.2017.12.004

[37] A. Wilhite, E. Fong, and S. Wilhite, "The influence of editorial decisions and the academic network on self-citations and journal impact factors," *Research Policy,* vol. 48, no. 6, pp. 1513–1522 (2019). [Online]. Available: https://doi.org/10.1016/j.respol.2019.03.003

[38] R. Van Noorden and D. S. Chawla, "Hundreds of extreme self-citing scientists revealed in new database," *Nature,* vol. 572, no. 7771, pp. 578–580 (2019). [Online]. Available: https://doi.org/10.1038/d41586-019-02479-7

[39] J. Fowler and D. Aksnes, "Does self-citation pay?" *Scientometrics,* vol. 72, no. 3, pp. 427–437 (2007). [Online]. Available: https://doi.org/10.1007/s11192-007-1777-2

[40] A. Wilhite and E. Fong, "Coercive citation in academic publishing," *Science*, vol. 335, no. 6068, pp. 542–543 (2012). [Online]. Available: https://doi.org/10.1126/science.1212540

[41] B. Martin, "Whither research integrity? Plagiarism, self-plagiarism and coercive citation in an age of research evaluatement," *Research Policy*, vol. 42, no. 5, pp. 1005–1014 (2013). [Online]. Available: https://doi.org/10.1016/j.respol.2013.03.011

[42] M. Haley, "On the inauspicious incentives of the scholar-level h-index: An economist's take on collusive and coercive citation," *Applied Economics Letters*, vol. 24, no. 2, pp. 85–89 (2017). [Online]. Available: https://doi.org/10.1080/13504851.2016.1164812

[43] R. Van Noorden, "Brazilian citation scheme outed," *Nature*, vol. 500, no. 7464, pp. 510–511 (2013). [Online]. Available: https://doi.org/10.1038/500510a

[44] R. Van Noorden, "Highly cited researcher banned from journal board for citation abuse," *Nature,* vol. 578, no. 7794, pp. 200–202 (2020). [Online]. Available: https://doi.org/10.1038/d41586-020-00335-7

[45] A. Grudniewicz, D. Moher, K. D. Cobey, G. L. Bryson, S. Cukier, K. Allen, C. Ardern, P. Balcom, L. Barros, M. Berger, M. Cirocco, M. C. Cobey, M. C. Lalu, L. Laupacis, and J. M. Shamsher, "Predatory journals: no definition, no defence," *Nature*, vol. 576, pp. 210–212 (2019).

[46] M. Biagioli, "Watch out for cheats in citation game," *Nature,* vol. 535, no. 7611, p. 201 (2016). [Online]. Available: https://doi.org/10.1038/535201a

[47] P. Nattavudh, Y. Riyanto, and J. Knetsch, "Lower-rated publications do lower academics' judgments of publication lists: Evidence from a survey experiment of economists," *Journal of Economic Psychology,* vol. 66, pp. 33–44 (Jun. 2018). [Online]. Available: https://doi.org/10.1016/j.joep.2018.04.003

[48] H. F. Moed, L. Colledge, J. Reedijk, F. Moya-Anegón, V. Guerrero-Bote, A. Plume, and M. Amin, "Citation-based metrics are appropriate tools in journal evaluation provided that they are accurate and used in an informed way," *Scientometrics*, vol. 92, no. 2, pp. 367–376 (2012). [Online]. Available: https://doi.org/10.1007/s11192-012-0679-8

[49] I. N. Sengupta, "Bibliometrics, informetrics, scientometrics and librametrics: An overview," *Libri,* vol. 42, no. 2, pp. 75–98 (1992). [Online]. Available: https://doi.org/10.1515/libr.1992.42.2.75

[50] L. Bornmann, R. Mutz, S. Hug, and H. D. Daniel, "A multilevel meta-analysis of studies reporting correlations between the h index and 37 different h index variants," *Journal of Informetrics,* vol. 5, no. 3, pp. 346–359 (2011). [Online]. Available: https://doi.org/10.1016/j.joi.2011.01.006

[51] L. Wildgaard, J. W. Schneider, and B. Larsen, "A review of the characteristics of 108 author-level bibliometric indicators," *Scientometrics*, vol. 101, no. 1, pp. 125–158 (2014). [Online]. Available: https://doi.org/10.1007/s11192-014-1423-3

[52] L. Waltman, "A review of the literature on citation impact indicators," *Journal of Informetrics,* vol. 10, no. 2, pp. 365–391 (2016). [Online]. Available: https://doi.org/10.1016/j.joi.2016.02.007

[53] L. Bornmann, R. Mutz, C. Neuhaus, and H. D. Daniel, "Citation counts for research evaluation: Standards of good practice for analyzing bibliometric data and presenting and interpreting results," *Ethics in Science and Environmental Politics*, vol. 8, no. 1, pp. 93–102 (2008). [Online]. Available: https://doi.org/10.3354/esep00084

[54] J. Qiu, R. Zhao, S. Yang, and K. Dong, Informetrics: Theory, Methods and Applications. Cham: Springer (2017).

[55] R. Rousseau, L. Egghe, and R. Guns, Becoming Metric-Wise: A Bibliometric Guide for Researchers. Duxford, UK: Chandos Publishing (2018).

[56] E. Roldan-Valadez, S. Y. Salazar-Ruiz, R. Ibarra-Contreras, and C. Rios, "Current concepts on bibliometrics: A brief review about impact factor, Eigenfactor score, CiteScore, SCImago Journal Rank, Source-Normalised Impact per Paper, h-index, and alternative metrics," *Irish Journal of Medical Science* (1971-), vol. 188, no. 3, pp. 939–951 (2019). [Online]. Available: https://doi.org/10.1007/s11845-018-1936-5

[57] A. Ahmi, Bibliometric Analysis for Beginners: A starter guide to begin with a bibliometric study using Scopus dataset and tools such as Microsoft Excel, Harzing's Publish or Perish and VOSviewer software (2021).

[58] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," *Journal of Business Research*, vol. 133, pp. 285–296 (2021).

[59] N. Bontis and A. Serenko, "A follow-up ranking of academic journals," *Journal of Knowledge Management,* vol. 13, no. 1, pp. 16–26 (2009). [Online]. Available: https://doi.org/10.1108/13673270910931134

[60] P. O. Seglen, "Why the impact factor of journals should not be used for evaluating research," *British Medical Journal,* vol. 314, pp. 498–502 (1997). [Online]. Available: https://doi.org/10.1136/bmj.314.7079.497

[61] P. O. Seglen, "Citation rates and journal impact factors are not suitable for evaluation of research," *Acta Orthopaedica Scandinavica*, vol. 69, no. 3, pp. 224–229 (1998). [Online]. Available: https://doi.org/10.3109/17453679809000920

[62] A. Oswald, "An examination of the reliability of prestigious scholarly journals: Evidence and implications for decision-makers," *Economica*, vol. 74, no. 293, pp. 21–31 (2007). [Online]. Available: https://doi.org/10.1111/j.1468-0335.2006.00575.x

[63] J. Bar-Ilan and G. Halevi, "Post retraction citations in context: A case study," *Scientometrics,* vol. 113, pp. 547–565 (2017). [Online]. Available: https://doi.org/10.1007/s11192-017-2242-0

[64] V. Larivière and C. R. Sugimoto, "The journal impact factor: A brief history, critique, and discussion of adverse effects," in Springer Handbook of Science and Technology Indicators, W. Glänzel, H. F. Moed, U. Schmoch, and M. Thelwall, Eds., Cham: Springer, pp. 3–24 (2019). [Online]. Available: https://doi.org/10.1007/978-3-030-02511-3_1

[65] M. Haley and M. McGee, "A parametric 'parent metric' approach for comparing maximum-normalized journal ranking metrics," *Journal of the Association for Information Science and Technology,* vol. 69, no. 1, pp. 172–176 (2018). [Online]. Available: https://doi.org/10.1002/asi.23908

[66] A. W. Harzing, Publish or Perish (2007). [Online]. Available: https://harzing.com/resources/publish-or-perish

[67] J. J. Heckman and S. Moktan, "Publishing and promotion in economics: The tyranny of the top five," *Journal of Economic Literature,* vol. 58, no. 2, pp. 419–470 (2020). [Online]. Available: https://doi.org/10.1257/jel.20191574

[68] A. Pudovkin and E. Garfield, "Rank-normalization impact factor: A way to compare journal performance across subject categories," in Proc. 67th Annu. Meeting of the American Society for Information Science and Technology, vol. 41, no. 1, pp. 507–515 (2004). [Online]. Available: https://doi.org/10.1002/meet.1450410159

[69] F. Radicchi, S. Fortunato, and C. Castellano, "Universality of citation distributions: Toward an objective measure of scientific impact," Proceedings of the National Academy of Sciences USA, vol. 105, no. 45, pp. 17268–17272 (2008). [Online]. Available: https://doi.org/10.1073/pnas.0806977105

[70] H. F. Moed, "Measuring contextual citation impact of scientific journals," *Journal of Informetrics,* vol. 4, no. 3, pp. 265–277 (2010). [Online]. Available: https://doi.org/10.1016/j.joi.2010.01.002

[71] L. Leydesdorff, P. Zhou, and L. Bornmann, "How can impact factors be normalized across fields of science? An evaluation in terms of percentile ranks and fractional counts," *Journal of the American Society for Information Science and Technology*, vol. 64, no. 1, pp. 96–107 (2013). [Online]. Available: https://doi.org/10.1002/asi.22765

[72] L. Waltman and N. J. van Eck, "Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison," *Scientometrics,* vol. 96, no. 3, pp. 699–716 (2013). [Online]. Available: https://doi.org/10.1007/s11192-012-0913-4

[73] S. Alonso, F. Cabrerizo, E. Herrera-Viedma, and F. Herrera, "hg-index: A new index to characterize the scientific output of researchers based on the h-and g-indices," *Scientometrics,* vol. 82, no. 2, pp. 391–400 (2010). [Online]. Available: https://doi.org/10.1007/s11192-009-0047-5

[74] R. S. Tol, "The Matthew effect defined and tested for the 100 most prolific economists," *Journal of the American Society for Information Science and Technology*, vol. 60, no. 2, pp. 420–426 (2009). [Online]. Available: https://doi.org/10.1002/asi.20968

[75] V. Larivière and Y. Gingras, "The impact factor's Matthew Effect: A natural experiment in bibliometrics," *Journal of the American Society for Information Science and Technology,* vol. 61, no. 2, pp. 424–427 (2010). [Online]. Available: https://doi.org/10.1002/asi.21232

[76] K. Drivas and D. Kremmydas, "The Matthew effect of a journal's ranking," *Research Policy,* vol. 49, no. 4, art. 103951 (2020). [Online]. Available: https://doi.org/10.1016/j.respol.2020.103951

[77] R. S. Tol, "The Matthew effect for cohorts of economists," *Journal of Informetrics*, vol. 7, no. 2, pp. 522–527 (2013). [Online]. Available: https://doi.org/10.1016/j.joi.2013.02.001

[78] S. Stremersch, I. Verniers, and P. C. Verhoef, "The quest for citations: Drivers of article impact," *Journal of Marketing*, vol. 71, no. 3, pp. 171–193 (2007). [Online]. Available: https://doi.org/10.1509/jmkg.71.3.

[79] M. Haley, "A simple paradigm for augmenting the Euclidean index to reflect journal impact and visibility," *Journal of the Association for Information Science and Technology,* vol. 71, no. 3, pp. 370–373 (2019b). [Online]. Available: https://doi.org/10.1002/asi.24224

[80] M. Haley, "Combining the weighted and unweighted Euclidean Indices: A graphical approach," *Scientometrics*, vol. 123, pp. 103–111 (2020). [Online]. Available: https://doi.org/10.1007/s11192-020-03368-x

[81] H. Piwowar, "Introduction altmetrics: What, why and where?" *Bulletin of the American Society for Information Science and Technology,* vol. 39, no. 4, pp. 8–9 (2013). [Online]. Available: https://doi.org/10.1002/bult.2013.1720390404

[82] L. Bornmann, "Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics," *Journal of Informetrics*, vol. 8, no. 4, pp. 895–903 (2014). [Online]. Available: https://doi.org/10.1016/j.joi.2014.09.005

[83] F. Didegah, T. D. Bowman, and K. Holmberg, "On the differences between citations and altmetrics: An investigation of factors driving altmetrics versus citations for Finnish articles," *Journal of the Association for Information Science and Technology*, vol. 69, no. 6, pp. 832–843 (2018). [Online]. Available: https://doi.org/10.1002/asi.23934