



# In Search of an Academic “Greatest Hits” Album: An Examination of Bibliometrics and Bibliometric Web Platforms

Christopher R. Marsicano<sup>1</sup> · Alexander R. K. Nichols<sup>2</sup>

Accepted: 4 October 2022 / Published online: 5 November 2022  
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

## Abstract

Despite the relative youth of bibliometric web platforms (Google Scholar was released in 2004), they play an increasingly significant role in the assessment of the impact of scholars and the research they produce. This scholarly essay provides a thorough review of the literature on bibliometric platforms, the extent to which they make available relevant manuscripts for inclusion in research, and their use for the assessment of scholarly work. We describe the metrics found on common bibliometric platforms, proposed metrics not commonly found in platforms, and how those metrics may differ based on scholar race and gender. We identify pitfalls of citation metrics present on bibliometric platforms. Finally, we identify areas for expansion of the research on bibliometric platforms and development of new metrics.

**Keywords** Bibliometric platforms · Google Scholar · Citation disparities

**JEL Classification** A29 · Y20 · Y80

Web-based bibliometric platforms like Google Scholar and Scopus are useful tools to ensure that scholars who “stand on the shoulders of giants” build upon and extend the work of the great thinkers before them. Platforms allow easy searching of extant literature, trace citation patterns, and speed up the process of research relative to the years before the internet. Bibliometric platforms also allow for comparisons of scholars and the ability to evaluate scholars for career advancement, such as tenure and promotion. Their usefulness for a variety of purposes has been the subject of countless scholarly works.

---

✉ Alexander R. K. Nichols  
Nichols.864@osu.edu

<sup>1</sup> Department of Educational Studies, Davidson College, Box 7124, Davidson, NC 28035-7124, USA

<sup>2</sup> Department of Educational Studies, The Ohio State University College of Education and Human Ecology, 1990 College Rd N, Columbus, OH 43210, USA

This scholarly essay takes the form of a literature review on such platforms, with a focus on the ability of bibliometric platforms to produce relevant literature, the availability and ubiquity of peer-reviewed research found using search functions in bibliometric platforms, and their use in assessing scholars and their work.<sup>1</sup> We also discuss the metrics commonly found on bibliometric profiles, relevant metrics not commonly found on those profiles, and race- and gender-based differences in citation patterns and bibliometric platform metrics. Finally, we identify areas in need of greater development and study related to bibliometric platforms and the metrics they use.

## The Efficacy of Bibliometric Platforms

Given the relative youth of bibliometric platforms (Google Scholar began beta testing in 2004), the extant literature on such platforms is relatively new. Most of the available literature on Google Scholar, Web of Science, Scopus, and others falls into two major categories. First, researchers assess the availability of relevant manuscripts for inclusion in literature reviews and other scholarly work on bibliometric platforms. Second, scholars have examined the use of such platforms for evaluation purposes. We summarize the literature below, using Google Scholar as our main focus, believing that the efficacy of other platforms like Web of Science and Scopus is evident in their comparisons with Google Scholar. All of the aforementioned platforms have their own advantages and disadvantages, which we cover below.

In the years following the advent of bibliometric sources, much of literature has examined the ability of platforms to provide relevant, peer-reviewed manuscripts for use in scholarship – with Google Scholar being of outsized interest to researchers. In an essay concerning the benefits and drawbacks of Google Scholar, Jacsó (2005) argues that the platform omits important information - mostly due to indexing difficulties the platform exhibited then that are no longer the case today. The essay, written just one year after the advent of Google Scholar, discusses the results of a side-by-side analysis tool comparing web search yields per search query from Google Scholar to other major platforms. Based on the findings from that analysis, it argues that one of the great benefits of the platform is its ability to comb the internet for important work from scholarly publishers, university presses, professional societies, government agencies, and preprint/reprint servers. Jacsó also argues that the greatest beneficiaries of the platform are those with access to publishers' archives – as Google Scholar collates results from many sources into one easy to understand search result. Jacsó, however, criticizes two major “features” of Google Scholar. First, he argues that Google Scholar is only as good as the print sources it is allowed to search – and the platform is less than forthcoming as to the universe of its sources. To use Jacsó's example, “Google is as secretive about its coverage as the

---

<sup>1</sup> To compile the literature for this piece, we felt it appropriate to use bibliometric platforms to identify relevant literature. Every piece cited in this work was identified through bibliographic searches at our university libraries or through a Google Scholar, Web of Science, or Scopus search.

North Korean government about the famine in the country” (Jacsó, 2005, p. 209). As a result, researchers and policymakers in search of scholarship on a given subject may find only a fraction of the total number of research works on a given topic. Second, Jacsó states that Google Scholar limits indexing to only a portion of each source, and it is unclear how often the platform updates its sources.

Building off of these concerns, Neuhaus et al. (2006) compared Google Scholar to 47 other databases. The authors tested Google Scholar for publication date, publication language bias, and upload frequency date. They searched for a random selection of articles in databases covering multiple disciplines including business, education, the humanities, natural sciences, medicine, and the social sciences. They then determined whether the same article titles were available through Google Scholar. They found a high amount of overlap in search results between the established databases and Google Scholar in single publisher, open access, and natural sciences and medicine databases. Google Scholar did not perform as well, however, when searching for articles available in the social sciences, humanities, and education. Despite disciplinary differences in article availability in Google Scholar, Neuhaus et al. do assess the platform’s presentation of hyperlinks to open access research articles to be a major benefit to the platform - provided that one understands English. The authors found a clear bias towards English-language publications.

Research into Google Scholar’s coverage of medical research articles and their applications bolster Neuhaus et al.’s (2006) findings. Medical doctors have long received encouragement to use Google Scholar for browsing and “serendipitous discovery” when examining emerging health research (Gehanno et al., 2013, p. 1). Conventional wisdom in the field, however, suggested that doctors should not use Google Scholar for systematic reviews – a form of analysis used to synthesize all existing evidence on a particular issue of interest such as the efficacy of a drug trial or a particular surgical procedure. Gehanno et al. (2013) examined whether Google Scholar alone could provide enough high-quality research outputs to build a necessary base of knowledge for a systematic review. After selecting 738 “gold standard” studies from established databases from 29 high-quality systematic reviews, the authors searched Google Scholar for each of the titles to see if Google Scholar could assist medical researchers in covering all of the relevant literature. While some of the bibliographic references provided by Google Scholar included major errors, Gehanno et al. (2013) found each and every one of the 738 articles – including five articles in the so-called “Gray Literature... written material that is not published commercially or is not generally accessible” in the database, leading to 100% coverage (Gehanno et al., 2013). Papers presented at the meetings of professional and academic associations offer an example of grey literature. The authors argue that Google Scholar coverage of high-quality research is greater than previously thought, and that researchers should use the platform for systematic reviews and meta-analyses.

Researchers have compared bibliometric platforms in their ability to search both for classic works as well as new, not-yet or never-going-to-be peer-reviewed grey/gray literature. Much of the literature examines the extent to which citation managers such as Web of Science, Google Scholar, and Scopus differ in their search capabilities (Martín-Martín et al., 2018a, b). Martín-Martín et al. (2018b) examined

the differences between all three bibliometric citation platforms. The authors argue that Google Scholar presents more sources than the other platforms in the sciences, but Martín-Martín et al. wanted to examine academic areas outside the sciences. To understand which citation platform is best for each academic subject area, the authors took a non-random sample of articles from each of the platforms and classified them into 252 categories of academic areas. They found significantly more citations on Google Scholar across more academic backgrounds.

This finding could simply be due to Google Scholar's more expansive base of sources for citation counts. Yet that explanation alone does not account for platform citation count differences. Work by Marsicano et al. (2022) bolsters and extends Martín-Martín et al.'s (2018b) findings. Comparing the number of citations for scholars with profiles and citation counts in both Google Scholar and Scopus to those only in Scopus, the authors found that the counts for those with profiles in both platforms were statistically significantly higher in both platforms than those with Scopus profiles only. The authors speculated that Google Scholar's openness and lack of a financial barrier to entry could lead researchers to use it more often for searching for literature, eventually leading to pieces on that platform having higher citation counts. Thus, having a profile in Google Scholar could increase the number of citations of each publication on all platforms.

In another study from many of the same authors of the Martín-Martín et al. (2018b) piece, researchers examined the extent to which Google Scholar, Web of Science, and Scopus provide access to "classic papers" within multiple fields (Martín-Martín et al., 2018a). The authors hypothesized that because Google Scholar includes more sources in its searches, it may be better positioned than Web of Science or Scopus to provide a more holistic view of the state of the literature. They find more coverage of articles on Google Scholar, and that using only sites like Web of Science and Scopus may limit literature searches.

Researchers have also examined whether bibliometric platforms provide differential results when searching for the so-called "grey literature," the articles, conference presentations, and other manuscripts not yet formally published by academic or other commercial publishers. When Haddaway et al. (2015) examined the differences in the prevalence of grey literature across multiple citation managers, they found little difference in search availability of works of grey literature. Marsicano et al. (2022) also note that some bibliometric platforms allow for the merging of grey literature pieces with the eventually published versions.

Comparisons have also examined core differences in platforms in terms of manuscript availability and citation metrics. Critics of Web of Science argue that the platform covers mostly Western European and North American titles, does not count book titles, provides differential coverage between disciplines, and has citation errors due to its Anglophone focus (making non-Western names difficult to catalog) and inconsistencies in the use of initials (Meho & Yang, 2007). For example, Meho and Yang (2007) compared the citation metrics of 25 library science scholars across three platforms – Scopus, Web of Science, and Google Scholar. The authors find a near 60% overlap of the citations of each article in Web of Science and Scopus, but that the number of citations for the faculty in the middle of a distribution of number of citations per faculty member change dramatically from Web of Science and

Scopus (Meho & Yang, 2007). The authors also find that Google Scholar's inclusion of conference proceedings and international, non-Anglophone journals expands the citation counts for researchers in fields dominated by conferences and internally-focused, peer-reviewed journals (Meho & Yang, 2007).

In general, the literature suggests bibliometric platforms are a useful tool for searching for relevant literature for research and provide an assessment of the scholarly output of a researcher. There are many advantages of such systems - certainly over the card catalogs of old. Such advantages include the ability to access gray literature pieces and the ease of identifying important scholarly works, regardless of whether researchers use Google Scholar, Scopus, or Web of Science. That said, Google Scholar uses a more expansive database and, therefore, provides larger citation counts and associated metrics.

## **Pitfalls of Citation Metrics in Bibliometric Platforms**

While some have argued for the use of bibliometric platforms to evaluate scholars during high stakes professional decisions, such as those related to tenure and promotion (e.g. Marsicano et al., 2022), others have offered full-throated critiques of their use (Jensenius et al., 2018). At their core, bibliometric platforms are counters of citations. Differential citation patterns in the literature based on race and gender, therefore, pose a perennial challenge not just to fields and disciplines, but also to bibliometric platforms and their ability to facilitate rigorous science and evaluation (Jensenius et al., 2018). Bibliometric platforms are only as good as the data they use, and differential citation patterns across races and genders could impact the citation metrics present on platforms like Google Scholar and Scopus.

## **Disparities by Gender**

While there are now more white women, women of color, and men of color college professors than ever before, white male faculty still dominate academia, both in numbers and (more pointedly) influence. Using citation counts as a proxy for “influence,” several studies show that race and gender play a conspicuous part in which and whose scholarship is worthy of citation. Looking at the gender citation gap specifically, Dion et al. (2018) assert that male scholars achieve a higher number of citations attributed to their work than their female counterparts in the same fields across multiple disciplines. By analyzing every article published across three political science journals and three social science methodology journals from 2007 to 2016, Dion et al. (2018) found that male authors tend to cite other men over women in their article bibliographies. Their work also suggests that this pattern persists even in journals with a majority of female authors. Although the proportion of women working within the social sciences had increased notably in the decade analyzed, Dion et al. (2018) uncovered no evidence, evincing a trend that women were being cited more frequently.

King et al. (2017) document a similar tendency of male scholars to bolster their own citation counts by self-citing. The authors compile and codify 1.5 million research papers in the scholarly database JSTOR published between 1779 and 2011 to conduct their study. Their subsequent analysis shows that men cited their own papers 56% more often than women, with that number ballooning to 70% in the most recent two decades. Women scholars also were also 10 percentage points less likely to cite their work at all.

Some have argued that the gender citation gap could close as women made up an increasingly large proportion of the field, representing a critical mass (Ferber, 1988; Ferber & Brün, 2011; Dion et al., 2018). King et al. (2017) call into question this body of work by correlating the gender composition of a field's authorships with the rate of self-citation. The fields with the lowest women's self-citation rates per authorship were history (22.5%) and classical studies (22.3%) whereas ecology and evolution (29.4%), sociology (32.9%), and molecular and cell biology (26.8%) had the highest women's self-citation rates per authorship. In the field of international relations, Maliniak et al. (2013) show that, after controlling for several factors, an article written by a woman receives four citations for every five citations of an article written by a man. Even when representation is not the issue at hand, Zhang et al. (2021) show that female scholars are more likely to ask research questions related to societal progress than their male counterparts. Perhaps as a result, female scholars' work is read more often, but cited less than male scholars' work (Zhang et al., 2021). These articles suggest that increasing representation alone will not necessarily increase the citation counts of a marginalized group.

## Disparities by Race

Like with the gender citation gap, scholarly citations also vary by race. For example, Chakravartty et al. (2018) examined the racial citation gap by coding and analyzing the racial composition of primary authors of both articles and citations in communication studies journals from 1990 to 2016. The researchers provided evidence suggesting that communication studies as an academic discipline have a dearth of scholars of color and that said scholars often received fewer citations from peers. While a robust body of work proves that the racial citation gap exists, few precisely measure how racial bias impacts the manner in which a scholarly community engages with the ideas of scholars of color. One such work by Bertolero et al. (2020) assesses the extent and drivers of racial imbalance in the reference lists of papers published in five top neuroscience journals over a 25-year timespan. Major findings from their paper include that neuroscience reference lists tend to favor papers with a white scholar as first and last author—and that the disparity stems from the citation practices of white authors. Of note, Bertolero et al. (2020) also showed that papers with scholar of color as first and last authors were 17.2% less than expected based on racial/ethnic probabilities in the pool of citable papers. Another key finding that mirrors the gender citation gap is that the racial citation gap is also increasing, despite increased diversification in the field of neuroscience.

Research by groups such Chakravartty et al. (2018) and Bertolero et al. (2020) point to an underlying conclusion: That citations are inherently political. The notion of citations as political draws its roots from Richard Delgado's (1984) article "The Imperial Scholar," which highlights the racially biased citation patterns of a small group of (white) civil rights scholars. Specifically, the group in question tended to cite themselves to the exclusion of their black peers, allowing them to lead scholarship on African American civil rights. This reflective piece engendered and empowered critical race theorists to explore the race citation gap as a measurable phenomenon rather than an enigmatic musing.

## Disparities by Race and Gender

The corpus of work that exists on the racial citation gap and gender citation gap often evaluates both disparities as separate issues, removing the intersectional dimension of race and gender. Of the numerous studies on faculty disparities (in pay, recognition, impact, etc.), only a handful differentiate among racial and gender lines (see Hur et al., 2017). Since women faculty predominantly identify as white, studies on gender disparities that fail to disaggregate by race are generally reporting on the experiences of white women (Fox Tree & Vaid, 2022). Furthermore, research on racial disparities in academia tends to focus on minoritized groups with little discussion on the gendered experiences of these groups (e.g., Dimmick & Callahan, 2021). Thus, discussion on gender that does not consider how gender intersects with race (and vice versa) in effect erases the experiences of women of color faculty. In response, Fox Tree and Vaid (2022) call for more intersectionality-oriented datasets and studies.

One such study by Hopkins et al. (2013) aims to provide such a dataset by examining disparities in publication patterns across gender and based on a survey of a random sample of authors. In particular, Hopkins et al. (2013) surveyed a random sample of 1065 authors who contributed a peer-reviewed journal article indexed in the Web of Science (WoS) in 2005 and at least one other article during the period of 2001–2004 in four academic disciplines, namely biochemistry, water resources, economics, and anthropology. They then mapped the demographic variables (i.e., race and gender) onto the career-related variables (rank, discipline, h-index) of the sampled authors. At every career level and in each academic discipline, women (especially Black and Hispanic women) authors published at a lower percentage than their male peers at their level and in their respective fields. A by-product of this finding also showed that women of color generally boasted lower h-indices than men of any race and white women in all four disciplines surveyed.

The extent to which white women, women of color, and men of color academics lag behind (or in some cases, supersede) their white male counterparts does vary by discipline. Merritt (2000) found only modest differences in logged citation counts between white male law professors and women and minority law professors. She conducted her exploratory study on 815 professors who began tenure-track positions at accredited U.S. law schools between 1986 and 1991 and who remained on the tenure track in fall 1998. After controlling for race and gender differences through



regression equations, the citation gap between white men and both white and minority women closed, with a substantive reduction in the gap for minority men. According to Merritt (2000), variations in citation rates generated more from differences in educational background, prestige of the institution at which a professor teaches, and teaching assignments than from gender and race in and of themselves. These same variables explain most of the shortfall for men of color faculty, although some differences remain between these men and their white male colleagues even after controlling for race and gender.

Merritt's (2000) analysis points to the prominence of the "Matthew Effect," which describes the snowballing advantage that accrues to scholars who are already successful, in academia. This term served as the foundation of the later coined "Matilda Effect," which describes the phenomenon in which women's contributions are undervalued, or attributed to men (Rossiter, 1993). Both show that structural inequities more so than innate biological differences account for the differential success of a particular scholar in any given field, with some (but not all) of these differences attributed to the topics chosen by White male scholars as opposed to their marginalized peers (see Kozlowski et al., 2022). Indeed, Milard and Tanguy (2018) show that differences in the quality of work are not necessarily to blame either. They reported that authors tend to cite people they know with men more likely to cite other men, and white authors more likely to cite other white authors — and that this behavior was due, in part, to a tendency to co-author papers with individuals of the same gender or race.

### **Disparities by Innovation and Originality**

Bibliometric platforms are also limited with respect to innovation and originality (Jensenius et al., 2018). A scholar producing truly original work may go unnoticed and uncited for some time before a rampant increase in citations. Jensenius et al. (2018) use the example of John Nash who received only 16 citations for his paper that proposed "Nash equilibrium" in the first five years after its publication. At the time of writing, the paper, "Equilibrium points in N-person games", has been cited 9512 times on Google Scholar. It takes time for young scholars or those proposing innovative ideas to gain traction and to see their work cited. In short, bibliometric platforms have difficulty in projecting long-term impact from short-term citation counts and metrics. This could be especially concerning for early career scholars who are assessed on their potential to produce quality research over a long career in hiring, tenure, and promotion decisions.

On the other end of the career development spectrum, long-tenured scholars with many papers that have gone uncited will not be noticed as such on bibliometric platforms. Most platforms do not calculate uncited works in their metrics. It is possible that scholars who produce work that does not make any impact as evidenced by citation would not be punished in the same way early career scholars might be. The stakes later in a career are lower and the incentives to produce impactful research less compelling when compared to early career faculty (Marsicano et al., 2022). This pattern is especially concerning for prospective and early career faculty of color and



women faculty. As Hofstra et al. (2020) warn, women, aspiring faculty of color and women of color are the most likely to craft novel connections in their research, but are simultaneously the least likely to be rewarded for their innovation. Hofstra et al. (2020) elucidate how race, gender, originality, and innovation intersect to create structural inequality, with much of this inequality stemming from existing metrics for interpreting bibliometric data.

## Existing Metrics Found in Bibliometric Platforms

Web bibliometric platforms generally include some basic citation metrics. Number of citations, while the least sophisticated of bibliometric counts, is among the most ubiquitous on web bibliometric platforms. Most platforms include several other more sophisticated measures. This section details those measures, followed by a section that discusses metrics, proposed in the literature but not often included in bibliometric sites.

Most bibliometric platforms include a scholar's (and in many cases a journal's) *h*-index. The *h*-index is a response to critics of basic citation counts that suggest that counts themselves are an assessment of quantity, but may not assess sustained impact or consistency. A "one-hit-wonder" scholar may produce one highly cited work and never produce an influential piece again. The *h*-index accounts for that possibility as a measure of quantity *and* impact over time.

The *h*-index is the largest of scholarly works that have received that same number of citations. For example, an author with an *h*-index of three would have three scholarly pieces that have been cited at least three times each. This formula presents a problem for highly cited authors that have a few "greatest hits" publications: it doesn't matter if all three of the scholar's publications have exactly three citations or if each of the three publications has over 100 citations – the *h*-index would be the same – three.

Google further attempted to ameliorate the problem of the "greatest hits" professor by introducing its own *i10* index in 2011. The *i10* index is simply the number of a scholar's publications that have earned at least 10 citations. A scholar with three articles that have at least 10 citations each would, therefore, have an *h*-index of 3 *and* an *i10*-index of 3. The *i10*-index would differentiate that scholar from one who has only nine citations - three each for three articles. In effect, the *i10*-index provides greater weight to those with a small number of "greatest hits." The *i10*-index does not completely ameliorate the plight of the "greatest hits" professor - a scholar with three articles, each of which is cited 100 times would have the same *h*-index and *i10*-index values – but it does identify that the scholar in question has more than a few citations for each work.

No discipline or field owns the *h*-index or the *i10*-index, and the *h*-index especially has been examined by scholars in the humanities (Baneyx, 2008), the social sciences (Altman, 2012; Burrows, 2012), medicine and the natural sciences (Bornmann & Daniel, 2005; Hirsch, 2005, 2007; van Raan, 2006). It has the extraordinary advantage of simplicity - it is easy to count to 10. The metrics are also

reliable; anyone who looks them up will get the same numbers (Jensenius et al., 2018). Yet, while both measures are helpful for within discipline comparisons, they have limitations in across discipline comparisons (Marsicano et al., 2022). Some disciplines – especially those that value conference proceedings and journal articles – provide scholars with more opportunities to publish than book-focused disciplines. More chances to publish most assuredly increase the number of chances to be cited – leading to higher *h*-index and *i10*-index values. For example, Marsicano et al. (2022) found that scholars in a discipline that focuses on journals, Chemistry, had mean *h*-index values four times of those from scholars in a “book” discipline, History.

Neither of these metrics commonly found in bibliometric platforms is without its critics. Given disciplinary differences in citation patterns that affect *h*-index and *i10*-index numbers, Harzing and Van der Wal (2009) argue that such metrics are not wholly objective and that a knowledge of disciplinary differences is necessary to compare metrics. Self-citation is also a problem for both metrics (Sandnes, 2020; Marsicano et al., 2022). Self-citation increases *h*-index values and *i10*-index values. Because self-citation differs across race and gender lines (King et al., 2017) the failure of *h*-index and *i10*-index values to assess self-citations naturally could lead to differentially lower index values for women and people of color. Lastly, because there are fewer women and people of color in the academy, and because women – especially Black and Hispanic women – publish less than their white male peers (Hopkins et al., 2013), there may be fewer chances for such scholars to cite and be cited. As such, both the *i10*- and *h*-indexes could exacerbate existing disparities.

Also, as evidenced from the “one-hit wonder” and “greatest hits” scholar examples from above, a scholar with a small number of extraordinarily highly cited pieces and a scholar with a large number of barely cited pieces would have similar *h*- and *i10*- index values (Bornmann & Daniel, 2009; Marsicano et al., 2022). We call this the “album” problem as, like album sales for an album with one major hit and an album with many songs with a moderate following are counted the same, scholars with differential quality of output could have the same index values. Most importantly, the ability to attain an accurate count of a scholar’s citations to calculate commonly found indices requires a bibliometric platform that gathers from expansive sources. The output of an index calculation is only as good as the data that go into that calculation.

## Proposed Metrics to Offset Concerns with Existing Metrics

Scholars and bibliometric platforms have attempted to offset concerns with the *h*- and *i10*-indices. Scopus, Google, and Web of Science all have quality assurance procedures to draw in appropriate sources, with Google Scholar being the most expansive of the three platforms (Marsicano et al., 2022). To deal with self-citation issues, Sandnes (2020) developed a simple calculation to assess the number of self-citations by a scholar based on self-citation data from over 100,000 published researchers in Google Scholar. In the Sandnes method, a scholar’s *h*-index squared divided by the total number of citations from all publications yields a value to predict self-citations.

Based on Sandnes's analysis of self-citation behaviors in the Google Scholar data, values over 0.35 indicate a high level of self-citation while test values below 0.2 signify a low level of self-citation. This calculation can help scholars and bibliometric platforms alike quickly discern how much self-citations are propagating onto the  $h$ - and  $i10$  indices of any given author.

Eghe (2006a, b) attempted to offset the album problem by introducing the  $g$ -index. The  $g$ -index holds that for a set of publications ranked in order by the number of citations received from largest number to smallest number, the index value is the largest number of articles,  $g$ , that have received  $g^2$  citations. For example, a  $g$ -index of 5 represents that the top 5 most-cited articles of a scholar have been cited an average of 5 times for a total of  $5^2$  (25) times. Using the  $g$ -index, a one-hit wonder with a piece with 91 citations and nine other pieces with one citation each would have an  $h$ -index of 1, an  $i10$ -index of 1, and a  $g$ -index of 10. A scholar with 10 pieces cited 10 times each would have  $h$ -,  $i10$ -, and  $g$ -index values of 10. The  $g$ -index, therefore, allows comparison of one-hit-wonder scholars to those with sustained, albeit moderate, success. In effect, it allows the scholarly equivalent of comparisons between Sir Mix-a-lot and Missy Elliott.<sup>2</sup>

Despite these advances, most bibliometric platforms include neither the  $g$ -index nor the Sandnes (2020) indicator on their platforms. Neither should be particularly hard to automate as both are calculable with raw citation counts. That said, neither metric offsets concerns about differential citation patterns based on race and gender. The Sandnes (2020) indicator does identify potential self-aggrandizing but does not guarantee it. There could be appropriate reasons for self-citation, especially for scholars trailblazing new areas of research for which there are few citations on which to draw or scholars showcasing how their previous research led to a long body of work in the same area. Given the lack of acknowledgement that bibliometric platforms give early career scholars or those proposing novel areas of work, self-citation may be necessary to improve overall visibility of early career faculty and those pursuing brand new ideas.

The Sandnes indicator and  $g$ -index do not solve the problems of uncited publications or lack of a clear measure of sustained impact. Think of the album problem. To use the album analogy, while the  $g$ -index places Manfred Mann and Bruce Springsteen<sup>3</sup> in the same league, it does not account for the sustained impact of the Boss over time. Marsicano et al. (2022) attempted to address these issues by suggesting a new metric, the  $u$ -index, and "percent" variations on the  $i10$ - and  $u$ - metrics. The  $u$ -metric is a simple sum of the number of articles of a scholar that go uncited. The  $u$ -percent metric is the percentage of a scholar's articles found in a bibliometric platform that go uncited five years after publication. Both metrics are designed to gauge

<sup>2</sup> Sir Mix-a-lot's "Baby Got Back" reached number 1 for five weeks in 1992. Missy Elliott has never had a number 1 hit, despite charting in the "Billboard Hot 100" 33 times, and in the top 10 on 9 occasions. Sir Mix-a-lot never had another song chart in the Top 10.

<sup>3</sup> Despite having 26 singles chart on the Billboard Hot 100 (with 12 in the top 10), Bruce Springsteen has never had a song hit number 1 as a performer. Manfred Mann's Earth Band's cover of Bruce Springsteen's "Blinded by the Light" topped the charts at number 1 in 1977. Manfred Mann's only hit was written by Bruce Springsteen.

the “quality of the quantity” of a scholar’s work (Marsicano et al., 2022). A low percentage of uncited publications and/or high number of uncited publications might suggest a lack of impactful research from the scholar. This lack of impact could be due to inequitable citation patterns by race or gender, a commitment to innovative work (the potential of which is not yet seen by a field), or it could simply be due to producing low-quality work. Marsicano et al. (2022) argue the *u*-metrics are best interpreted as an indicator of low-quality work. That said, racial and gender inequities and innovative work could theoretically cause quality papers to go uncited. Given the time period of five years proposed by Marsicano et al. (2022), however, it is unlikely that work that has even a small amount of impact would go completely uncited. Even Jensenius et al. example of delayed citation of impactful work - John Nash’s piece – had citations within 5 years of publication. Given that scholars of color and women cite work from other scholars of color and women (Dion et al., 2018), it is also unlikely that a lack of citations leading to the *u*-metrics could be wholly explained by gender and race differences. While minority status may certainly be a contributing factor to a lack of citations, Merritt’s (2000) analysis of law school professor citations might suggest other causes for lack of citable work.

The *i10*-percent indicator is the percentage of a scholar’s total number of articles that have at least 10 citations. Put another way, it is calculated by dividing a researcher’s Google Scholar *i10*-index value by the total number of articles found in the bibliometric platform. Marsicano et al. (2022) argue that the greater the *i10*-percent indicator, the greater the impact of a faculty member’s body of publications. Put in the language of the album problem, the greater the *i10*-percent index, the greater the likelihood that a scholar’s work goes platinum with each release. A sustained impact, as shown through a high *i10*-percent value, shows that a scholar’s work is of value to other scholars. That said, the metric alone does not solve the bibliometric problem of differential citation rates across disciplines, race, and gender. While it does show sustained impact, the scholar could benefit from self-citation and from citation by peers within the same demographic group. While the metrics proposed by Marsicano et al. (2022) do provide some answers to complicated problems related to bibliometric platforms and metrics, they are not a panacea.

## Areas Requiring Further Study and Metrics Development

Bibliometric platforms as of now focus generally on fairly basic metrics – the number of citations, the *i10*- and *h*-indices, etc. Research on these metrics has focused on eliminating ways to game the metrics or offset serious concerns around gender- and race-based differential citation patterns. There are, however, several areas in the development of bibliometric platforms that have not received thorough scholarly treatment.

## Interdisciplinary Research

According to Dogan and Pahre (1990), research confined to siloed fields of study (or “disciplines”) is less likely to yield new knowledge than research that cuts across

disciplines. Therefore, emergent disciplines—particularly, the modern sciences—now necessitate researchers from multiple disciplines to effectively address principal research questions (Hessels & van Lente, 2008; Raasch et al., 2013). Such trends have led scholars to argue for a greater integration of interdisciplinary methods in the research process (see Chubin, 1976; Nissani, 1997; Metzger & Zare, 1999; Forman & Markus, 2005; Raasch et al., 2013). Van Noorden (2015) attempts to identify and map interdisciplinary work by counting the number of research papers that cite work outside of their traditional disciplines. His findings indicate an upsurge in the fraction of papers with references to work in other disciplines in both the natural and social sciences and a decline in the fraction that point to another specialty in the same discipline. Despite a rise in interdisciplinary work, there is no clear metric within bibliometric platforms to identify scholarly interdisciplinarity, despite the fact that journals such as *Nature*, *Science*, and *PLOS One* have an interdisciplinary focus and place among the top most-cited journals annually. By developing metrics or indicators of interdisciplinary research, bibliometric platforms may incentivize greater collaborations across disciplines, thereby assuaging Dogan and Pahre's (1990) concerns. The need is clear for a transdisciplinary metric for inclusion in bibliometric platforms.

Transdisciplinary metrics could provide an indication of the extent to which a scholar's published work impacts the work of scholars in fields of study other than that of the focal scholar. We propose the development of a transdisciplinary metric that pertains to usefulness of a scholar's work to the work of scholars in different academic disciplines or fields of study other than that of the focal scholar. How frequently is the scholarship of a faculty candidate who is a political scientist cited by sociologists? How frequently is the scholarship of a chemistry faculty candidate cited by biologists? Bibliometric platforms could compute the proportion of the total number of citations to a scholar's publications that are cited by publications in academic disciplines or fields of study other than that of the faculty candidate. Bibliometric platforms generally already classify journals into disciplines. For example, Google Scholar already separates journals into 8 categories with titles like "Humanities, Literature, and Arts" and "Physics and Mathematics." Scopus's journal rankings system has an even more expansive and specific list of disciplines and fields numbering well over 100. To calculate the metric, a bibliometric platform that allows for scholars to create a profile (e.g. Google Scholar, Scopus, etc), would ask a scholar to identify their main discipline from a pre-populated list of disciplines and fields. The platform would then determine the number of citations in journals that do not match the disciplinary selection of the scholar. The platform would then compute the percentage of the total number of citations to the selected publications of a faculty candidate that are cited by publications in academic disciplines or fields of study other than the discipline chosen by the faculty candidate.

## International Research

Helms (2015) reports the findings of a study conducted by the American Council of Education on the contents of 91 tenure and promotion codes of primarily doctoral

and masters granting institutions. Of the 91 codes, 47 of them specify an international scholarly reputation as a criterion for tenure and promotion. Bibliometric platforms have the databases necessary to compute and present a metric that could serve as a proxy for an international scholarly reputation.

We propose bibliometric platforms to consider developing an estimate of the usefulness of the published work of the faculty candidate to the published works of scholars in other countries. For example, how frequently is the scholarship of a faculty candidate holding an academic appointment at a US-based college or university cited by scholars in the United Kingdom, The Netherlands, India, or South Africa? This metric is computed as the percentage of the total number of citations to the selected publications of a faculty candidate that are cited in published works of scholars in other nations. Given institutional affiliations are present in bibliometric platforms like Google Scholar and Scopus, it would not be difficult for a platform to compute the number of citations of a piece written by a scholar working at Queens University of Charlotte in the United States from scholars at Queens University in Canada and Queens University-Belfast in Northern Ireland. This international metric would provide researchers with a better understanding of the reach of an article beyond the borders of the country in which it originated.

## Conclusion

This article examines the extant literature on bibliometric platforms and the metrics they use. In it, we discuss the extent to which bibliometric platforms make available manuscripts needed for research, and the use of such platforms for the assessment of scholarly work. We examined the extent to which citation counts and other metrics are available on bibliometric platforms and how those metrics differ by author race and gender. We also identified critiques of various bibliometric measures and discussed the alternative metrics designed to offset those critiques. Lastly, we identified areas in need of further development and research – specifically international and interdisciplinary research – and made recommendations as to how bibliometric platforms could assess those areas. Further research should examine the research topic choices of women scholars and scholars of color as compared to their white male colleagues and how this difference in scholarly focus exacerbates disparities across race and gender. Furthermore, researchers should dive into self-citation behaviors and their impact on various metrics; self-citing shows a longstanding body of work in a research area, but also has the potential to “pad the stats” of a scholar. Researchers should examine the question, “how much is too much?”

In sum, this manuscript has four major takeaways. First, bibliometric platforms are useful tools in research. They can help scholars quickly identify important works and enable them to push forward the frontiers of science by extending previous work, thus speeding up the advancement of science. Second, they can be a valuable tool for assessing the impact of a scholar. Third, bibliometric platforms are only as good as the data they present. This means that pre-existing differential citation patterns, based on race and gender, are not offset in the metrics used on bibliometric platforms. Alternative metrics may be needed to ameliorate this issue. Lastly,

given the relatively infancy of web-based bibliometric platforms, there is still room to grow. New metrics must be developed to assess the reach of research across disciplines and national borders – and new scholarly work must evaluate those metrics.

**Data Availability** Not applicable for this work.

**Code Availability** Not applicable for this work.

## Declarations

**Conflict of Interest** The authors have no conflicts of interest to disclose.

## References

- Altman, D. (2012). Where is knowledge generated? On the productivity and impact of political science departments in Latin America. *European Political Science*, 11(1), 71–87. <https://doi.org/10.1057/eps.2010.82>
- Baneyx, A. (2008). “Publish or perish” as citation metrics used to analyze scientific output in the humanities: International case studies in economics, geography, social sciences, philosophy, and history. *Archivum Immunologiae et Therapiae Experimentalis*, 56(6), 363–371. <https://doi.org/10.1007/s00005-008-0043-0>
- Bertolero, M. A., Dworkin, J. D., David, S. U., Lloreda, C. L., Srivastava, P., Stiso, J., Zhou, D., Dzirasa, K., Fair, D. A., Kaczurkin, A. N., Marlin, B. J., Shohamy, D., Uddin, L. Q., Zurn, P., & Bassett, D. S. (2020). Racial and ethnic imbalance in neuroscience reference lists and intersections with gender. *BioRxiv*. <https://doi.org/10.1101/2020.10.12.336230>
- Bornmann, L., & Daniel, H. D. (2005). Does the h-index for ranking of scientists really work? *Scientometrics*, 65(3), 391–392. <https://doi.org/10.1007/s11192-005-0281-4>
- Bornmann, L., & Daniel, H.-D. (2009). The state of h index research. *EMBO Reports*, 10(1), 2–6. <https://doi.org/10.1038/embor.2008.233>
- Burrows, R. (2012). Living with the h-index? Metric assemblages in the contemporary academy. *The Sociological Review*, 60(2), 355–372. <https://doi.org/10.1111/j.1467-954X.2012.02077.x>
- Chakravartty, P., Kuo, R., Grubbs, V., & McIlwain, C. (2018). #CommunicationSoWhite. *Journal of Communication*, 68(2), 254–266. <https://doi.org/10.1093/joc/jqy003>
- Chubin, D. E. (1976). State of the field the conceptualization of scientific specialties. *The Sociological Quarterly*, 17(4), 448–476. <https://www.tandfonline.com/doi/abs/10.1111/j.1533-8525.1976.tb01715.x>
- Delgado, R. (1984). The imperial scholar: Reflections on a review of civil rights literature. *University of Pennsylvania Law Review*, 132(3), 561–578. <https://doi.org/10.2307/3311882>
- Dimmick A. A., Callahan J. L. (2021). Racial and ethnic diversity among clinical psychology doctoral students applying for internship. *Training and Education in Professional Psychology*. Advance online publication. <https://doi.org/10.1037/tep0000382>
- Dion, L. M., Sumner, J. L., & Mitchell, S. M. (2018). Gendered citation patterns across political science and social science methodology fields. *Political Analysis*, 26(3), 312–327. <https://doi.org/10.1017/pan.2018.12>
- Dogan, M., & Pahre, R. (1990). Scholarly reputation and obsolescence in the social-sciences: Innovation as a team sport. *International Social Science Journal*, 42(1), 417–427.
- Egghe, L. (2006a). How to improve the h-index. *The Scientist*, 20(3), 15.
- Egghe, L. (2006b). Theory and practise of the g-index. *Scientometrics*, 69(1), 131–152. <https://doi.org/10.1007/s11192-006-0144-7>
- Ferber, M. A. (1988). Citations and networking. *Gender and Society*, 2(1), 82–89. <https://doi.org/10.1177/089124388002001006>
- Ferber, M. A., & Brün, M. (2011). The gender gap in citations: Does it persist? *Feminist Economics*, 17(1), 151–158. <https://doi.org/10.1080/13545701.2010.541857>



- Forman, J., & Markus, M. L. (2005). Research on collaboration, business communication, and technology. *Journal of Business Communication*, 42(1), 78–102. <https://doi.org/10.1177/0021943604271958>
- Fox Tree, J. E., & Vaid, J. (2022). Why so few, still? Challenges to attracting, advancing, and keeping women faculty of color in academia. *Frontiers in Sociology*, 6, 92–198. <https://doi.org/10.3389/fsoc.2021.792198>
- Gehanno, J. F., Rollin, L., & Darmoni, S. (2013). Is the coverage of Google Scholar enough to be used alone for systematic reviews. *BMC Medical Informatics and Decision Making*, 13(7), 1–5. <https://doi.org/10.1186/1472-6947-13-7>
- Haddaway, N. R., Collins, A. M., Coughlin, D., & Kirk, S. (2015). The role of Google Scholar in evidence reviews and its applicability to grey literature searching. *PLoS One*, 10(9), e0138237. <https://doi.org/10.1371/journal.pone.0138237>
- Harzing, A.-W., & van der Wal, R. (2009). A Google Scholar h-index for journals: An alternative metric to measure journal impact in economics and business. *Journal of the American Society for Information Science and Technology*, 60(1), 41–46. <https://doi.org/10.1002/asi.20953>
- Helms, R. M. (2015). *Internationalizing the tenure code: Policies to promote a globally focused faculty*. American Council on Education.
- Hessels, L. K., & van Lente, H. (2008). Re-thinking new knowledge production: A literature review and a research agenda. *Research Policy*, 37(4), 740–760. <https://doi.org/10.1016/j.respol.2008.01.008>
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), 16569–16572. <https://doi.org/10.1073/pnas.0507655102>
- Hirsch, J. E. (2007). Does the h index have predictive power? *Proceedings of the National Academy of Sciences*, 104(49), 19193–19198. <https://doi.org/10.1073/pnas.0707962104>
- Hofstra, B., Kulkarni, V. V., Munoz-Najar Galvez, S., He, B., Jurafsky, D., & McFarland, D. A. (2020). The diversity-innovation paradox in science. *Proceedings of the National Academy of Sciences*, 117(17), 9284–9291. <https://doi.org/10.1073/pnas.1915378117>
- Hopkins, A. L., Jawitz, J. W., Mccarty, C., Goldman, A., & Basu, N. B. (2013). Disparities in publication patterns by gender, race and ethnicity based on a survey of a random sample of authors. *Scientometrics*, 96(2), 515–534. <https://doi.org/10.1007/s11192-012-0893-4>
- Hur, H., Andalib, M. A., Maurer, J. A., Hawley, J. D., & Ghaffarzadegan, N. (2017). Recent trends in the U.S. behavioral and social sciences research (BSSR) workforce. *PLoS One*, 12(2), e0170887. <https://doi.org/10.1371/journal.pone.0170887>
- Jacsó, P. (2005). Google Scholar: The pros and the cons. *Online Information Review*, 29(2), 208–214. <https://doi.org/10.1108/14684520510598066>
- Jensenius, F. R., Htun, M., Samuels, D. J., Singer, D. A., Lawrence, A., & Chwe, M. (2018). The benefits and pitfalls of Google Scholar. *Political Science and Politics*, 51(4), 1–5. <https://doi.org/10.1017/s104909651800094x>
- King, M. M., Bergstrom, C. T., Correll, S. J., Jacquet, J., & West, J. D. (2017). Men set their own cites high: Gender and self-citation across fields and over time. *Socius: Sociological Research for a Dynamic World*, 3, 1–22. <https://doi.org/10.1177/2378023117738903>
- Kozlowski, D., Larivière, V., Sugimoto, C. R., & Monroe-White, T. (2022). Intersectional inequalities in science. *Proceedings of the National Academy of Sciences*, 119(2), 1–8. <https://doi.org/10.1073/pnas.2113067119>
- Maliniak, D., Powers, R., & Walter, B. F. (2013). The gender citation gap in international relations. *International Organization*, 67(4), 889–922. <https://doi.org/10.1017/S0020818313000209>
- Marsicano, C., Braxton, J. M., & Nichols, A. R. K. (2022). The use of Google Scholar for tenure and promotion decisions. *Innovative Higher Education*, 47(4), 639–660. <https://doi.org/10.1007/s10755-022-09592-y>
- Martín-Martín, A., Orduna-Malea, E., & Delgado López-Cózar, E. (2018a). Coverage of highly-cited documents in Google Scholar, web of science, and Scopus: A multidisciplinary comparison. *Scientometrics*, 116(3), 2175–2188. <https://doi.org/10.1007/s11192-018-2820-9>
- Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & Delgado López-Cózar, E. (2018b). Google Scholar, web of science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, 12(4), 1160–1177. <https://doi.org/10.1016/j.joi.2018.09.002>
- Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58(13), 2105–2125. <https://doi.org/10.1002/asi.20677>

- Merritt, D. J. (2000). Scholarly influence in a diverse legal academy: Race, sex, and citation counts. *The Journal of Legal Studies*, 29(1), 345–368. <https://doi.org/10.1086/468076>
- Metzger, N., & Zare, R. N. (1999). Interdisciplinary research: From belief to reality. *Science*, 283(5402), 642–643. <https://doi.org/10.1126/science.283.5402.642>
- Milard, B., & Tanguy, L. J. (2018). Citations in scientific texts: Do social relations matter? *Journal of the Association for Information Science and Technology*, 69(11), 1380–1395. <https://doi.org/10.1002/asi.24061>
- Neuhaus, C., Neuhaus, E., Asher, A., & Wrede, C. (2006). The depth and breadth of Google Scholar: An empirical study. *Portal: Libraries and the Academy*, 6(2), 127–141. <https://doi.org/10.1353/pla.2006.0026>
- Nissani, M. (1997). Ten cheers for interdisciplinarity: The case for interdisciplinary knowledge and research. *The Social Science Journal*, 34(2), 14–37. [https://doi.org/10.1016/S0362-3319\(97\)90051-3](https://doi.org/10.1016/S0362-3319(97)90051-3)
- Raasch, C., Lee, V., Spaeth, S., & Herstatt, C. (2013). The rise and fall of interdisciplinary research: The case of open source innovation. *Research Policy*, 42(5), 1138–1151. <https://doi.org/10.1016/j.respol.2013.01.010>
- Rossiter, M. W. (1993). The Matthew Matilda effect in science. *Social Studies of Science*, 23(2), 325–341. <https://doi.org/10.1177/030631293023002004>
- Sandnes, F. E. (2020). A simple back-of-the-envelope test for self-citations using Google Scholar author profiles. *Scientometrics*, 124(2), 1685–1689. <https://doi.org/10.1007/s11192-020-03521-6>
- van Noorden, R. (2015). Interdisciplinary research by the numbers: An analysis reveals the extent and impact of research that bridges disciplines. *Nature*, 525(7569), 306–307. <https://doi.org/10.1038/525306a>
- van Raan, A. F. J. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, 67(3), 491–502. <https://doi.org/10.1556/Scient.67.2006.3.10>
- Zhang, L., Sivertsen, G., Du, H., Huang, Y., & Glänzel, W. (2021). Gender differences in the aims and impacts of research. *Scientometrics*, 126(11), 8861–8886. <https://doi.org/10.1007/s11192-021-04171-y>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

**Christopher R. Marsicano** PhD, is an Assistant Professor of Educational Studies and Public Policy and Founding Director of the College Crisis Initiative (C2i) at Davidson College. His research examines public policy and postsecondary education with a special focus on how higher education institutions behave as political actors to impact the policymaking process. As director of C2i, his expanded research portfolio includes working on the public health issues surrounding higher education innovation in the COVID-19 crisis.

**Alexander R. K. Nichols** is a PhD student in the Higher Education and Student Affairs program at The Ohio State University College of Education and Human Ecology. His research examines equity and inclusion practices, policies, and programs geared towards marginalized populations at postsecondary institutions (specifically predominantly White institutions).