The Benefits and Pitfalls of Google Scholar

Francesca R. Jensenius,† *University of Oslo*
Mala Htun, *University of New Mexico*
David J. Samuels, *University of Minnesota*
David A. Singer, *Massachusetts Institute of Technology*
Adria Lawrence, *Johns Hopkins University*
Michael Chwe, *University of California, Los Angeles*

ABSTRACT Google Scholar (GS) is an important tool that faculty, administrators, and external reviewers use to evaluate the scholarly impact of candidates for jobs, tenure, and promotion. This article highlights both the benefits of GS—including the reliability and consistency of its citation counts and its platform for disseminating scholarship and facilitating networking—and its pitfalls. GS has biases because citation is a social and political process that disadvantages certain groups, including women, younger scholars, scholars in smaller research communities, and scholars opting for risky and innovative work. GS counts also reflect practices of strategic citation that exacerbate existing hierarchies and inequalities. As a result, it is imperative that political scientists incorporate other data sources, especially independent scholarly judgment, when making decisions that are crucial for careers. External reviewers have a unique obligation to offer a reasoned, rigorous, and qualitative assessment of a scholar's contributions and therefore should not use GS.

When political scientists serve on hiring committees, evaluate candidates for tenure and promotion, and write letters of recommendation, we often are asked to assess a scholar's "impact." We provide our subjective appraisal of candidates' work and compare them against their peers. As they produce these sensitive and critical assessments, many political scientists rely on Google Scholar (GS) and other citation-count tools.

Francesca R. Jensenius *is associate professor of political science at the University of Oslo and senior research fellow at the Norwegian Institute of International Affairs. She can be reached at f.r.jensenius@stv.uio.no.*
Mala Htun *is professor of political science at the University of New Mexico. She can be reached at malahtun@gmail.com.*
David J. Samuels *is Distinguished McKnight University Professor of Political Science at the University of Minnesota. He can be reached at dsamuels@umn.edu.*
David A. Singer *is associate professor of political science at Massachusetts Institute of Technology. He can be reached at dasinger@mit.edu.*
Adria Lawrence *is the Aronson Associate Professor of International Studies and Political Science at Johns Hopkins University. She can be reached at adria.lawrence@jhu.edu.*
Michael Chwe *is professor of political science at the University of California, Los Angeles. He can be reached at michael@chwe.net.*
The † symbol indicates that the authors' names are in certified random order, as described in Ray and Robson (2018).

GS is readily available and requires no registration or subscription. Writers of tenure and promotion evaluation letters who, until the early 2000s engaged in mostly qualitative analysis of a candidate's file, now routinely refer to GS citation counts. Many political scientists have set up GS profile pages, which conveniently list all of a scholar's published (and often unpublished) work, along with citation counts for each work and summary statistics including total citation counts by year. Many scholars mention their citation count on their CVs, and some departments require faculty to report it in their tenure files.[1]

Some believe that citation counts are more objective than individual opinions about impact. Yet no objective metric of impact exists—of articles, books, journals, or individual scholars. Every metric contains built-in biases. Meanwhile, GS continues to be the discipline's *de facto* standard for assessment of scholarly impact despite a spirited discussion—in blogs, journals, and conferences—of its drawbacks (Hendrix 2015; Jascó 2005; Maliniak, Powers, and Walter 2013; Nexon 2016; Samuels 2011; 2013). Because political scientists seek unbiased measures to advance empirical arguments, it is especially odd to use a measure with so many obvious flaws to evaluate ourselves (Reiter 2016).

The purpose of this article is to amplify the discipline-wide discussion about evaluation criteria by reckoning with the benefits

doi:10.1017/S104909651800094X
https://doi.org/10.1017/S104909651800094X Published online by Cambridge University Press

and the pitfalls of using GS to assess scholarly impact.[2] Whereas the advantages we identify pertain primarily to GS, the disadvantages apply to almost all citation counts. The strengths of GS include incentives for quality, visibility, and open access; provision of a platform for networking; and reliability and consistency of its citation counts. However, GS has biases because citation is a social and political process that disadvantages certain groups, including women, younger scholars, scholars in smaller research communities, and scholars opting for risky and innovative over incremental work. GS counts also reflect practices of strategic citation that exacerbate existing hierarchies and inequalities.

The strengths of GS include incentives for quality, visibility, and open access; provision of a platform for networking; and reliability and consistency of its citation counts. However, GS has biases because citation is a social and political process that disadvantages certain groups, including women, younger scholars, scholars in smaller research communities, and scholars opting for risky and innovative over incremental work.

Many continue to use GS daily, which makes it imperative that we incorporate other data sources—especially independent scholarly judgment—when making decisions that are crucial for careers and lives. Although GS is useful for visibility and scholarly communication, we recommend against its use by external reviewers for tenure and promotion. GS counts should not have a double impact by figuring into both departmental *and* external evaluations. Instead, the unique role and obligation of expert external reviewers is to offer a reasoned, rigorous, and qualitative assessment of a scholar's contributions.

THE RISE OF GOOGLE SCHOLAR AND HOW IT WORKS

Google Scholar (GS) (www.scholar.google.com) was created in 2004 as a search engine for academics. Like Google's general search engine, it generates results based on the strength of the link between search terms and how often and how recently a work has been cited. GS indexes virtually everything available on the web in any language, including journal articles; academic books and book chapters; and non-peer-reviewed material such as conference papers, working papers, theses, and dissertations.[3]

GS also ranks the top 20 journals in a discipline. At the top of the GS home page, the "Metrics" button provides links to rankings by discipline and language, using an "H5 index"—that is, the number X articles that have at least X citations in the previous five years. Inexplicably, GS does not include international relations (IR) journals in its "top publications—political science" list.

Since 2012, Google has allowed scholars to create editable GS "profile pages" that it populates automatically with links to materials an author has written that are archived online. The profile page indicates how many times each item has been cited.

ADVANTAGES OF GOOGLE SCHOLAR

The advantages of GS stem primarily from its ease of use. A GS profile page provides a quick and convenient overview of a scholar's publications, rank-ordered by the number of citations of each publication. One can click on the hyperlinks of each publication to view abstracts and gain access to publicly available articles. Even articles that are stuck behind paywalls become more accessible because the program regularly "harvests" open-source versions of the articles from other websites. GS's benefits include incentives for quality and visibility, academic coordination and open access, and consistency in research evaluation.

Incentives for Quality and Visibility

Academics are under never-ending pressure to "publish or perish." In the United States, the tenure system puts a constant strain on faculty, particularly junior faculty. In the United Kingdom, this pressure was institutionalized through the Research Excellence Framework system, which makes public funding to universities contingent on publications and "impact."[4] The European Research Council requires all grant applicants to provide information about their "track record," including the number of publications and citations. In this context, easy access to GS may provide incentives for scholars to emphasize quality over quantity. They may want to produce better publications that are cited more instead of more publications with fewer citations. Focusing on citation counts also may inspire authors to publicize their work more, such as through social media.

Academic Coordination and Open Access to Research

GS may facilitate the dissemination of ideas and intellectual networking. A GS search on a keyword can expose a reader to a new scholar, and the reader can easily see the scholar's other work through their GS profile. The citation count shows which of the author's articles and books are most popular. In addition, GS's automatic email notifications can encourage intellectual networking. Interested readers can sign up to receive automatic emails when scholars post new work, and scholars can be notified when somebody cites them—often at the working-paper stage.

Our own experiences suggest that GS nudges scholars toward open-access practices (Hendrix 2016). When people see the convenient links to open-source versions of others' work, they may make their own work more publicly available via their personal websites or on academic networking sites (e.g., SSRN, Researchgate, and Academia.edu).

Consistency in Research Evaluation

The GS profile, with its full list of academic production, citation counts, and H-index, provides straightforward measures of scholarly quality and impact. These measures are highly reliable (i.e., anyone looking them up gets the same value) and consistent (i.e., the same coding scheme applies to all scholars). Given the prevalence of biases against women and people of color in the academy (Ginther et al. 2011; Milkman, Akinola, and Chugh 2012; Moss-Racusin et al. 2012), these measures offer grounds to challenge unfounded perceptions and prejudices.

Finally, the easy availability of citation counts can help scholars obtain credit for books and articles that are not published in the

https://doi.org/10.1017/S104909651800094X Published online by Cambridge University Press

most famous outlets but nevertheless are of intellectual importance. Citation counts can inform evaluators how the work has affected the field, independent of the publisher's prestige.

DISADVANTAGES OF GOOGLE SCHOLAR

What explains the variation in citation counts across scholars and their scholarly work? Like all data, GS citation counts are produced through social and political processes. Most would agree that good citation practices involve acknowledging prior work that helped generate ideas and explaining how one's claims fit into a literature. However, recent studies show that these commonsense citation practices disadvantage pioneering scholars with bold ideas, early-career scholars, scholars in smaller research communities, women, and solo authors. Furthermore, many cite works for strategic rather than principled reasons. As a result, citation counts do not provide an accurate assessment of scholarly impact and may exacerbate existing social hierarchies and inequalities.[5]

When people see the convenient links to open-source versions of others' work, they may make their own work more publicly available via their personal websites or on academic networking sites (e.g., SSRN, Researchgate, and Academia.edu).

Originality and Innovation

GS counts are biased toward incremental work and away from boldness and innovation. Highly original work that does not fit neatly into an existing literature might establish a new research agenda and expand interest in the topic, but its impact will not be visible in citation counts for many years. According to GS, John Nash's foundational paper defining Nash equilibrium received only 16 citations in the first five years after publication.

Another example is Stathis Kalyvas's 1999 article, "Wanton and Senseless? The Logic of Massacres in Algeria," which was cited only 20 times (excluding author citations and citations from drafts of the same article) in the four years after its publication—despite winning the Luebbert Award in 2000. However, its early citations came from a small group of colleagues who established a new agenda in the field of civil war and ethnic conflict, including Elisabeth Wood, John Mueller, David Laitin, and James Fearon. Kalyvas's (2006) book on civil wars currently has more than 3,000 citations, an impact impossible to predict from the 20 citations of his 1999 article four years after its publication.

In general, the number of citations that an article or book receives in the five or so years after publication reveals little about its long-term impact. Wang, Song, and Barabasi (2013) reviewed a sample of physics papers and found that having 50 citations in the first five years after publication was not associated with more citations after 20 years. In fact, papers with the most citations in 30 years tended to have relatively few citations early on. Stephan, Veugelers, and Wang (2017) examined 660,000 research articles in the Web of Science database and found that highly original papers were less likely to be highly cited within three years of publication but more likely to be highly cited three or more years after publication.

Early-Career Scholars

Because hiring and promotion decisions occur fairly early in a scholar's career, citation counts may be more consequential precisely when they are least informative. Papers and books that are cited many years after publication are arguably more important than papers and books cited only shortly after publication. The time-lag problem thus poses particular challenges in evaluating younger scholars.

Using GS to evaluate early-career scholars creates perverse incentives. From a numerical standpoint, it is better to publish incremental work on topics in which there is a large, active subgroup of scholars who cite one another than it is to open up a new field of research. Short-term, citation-centered evaluations discourage boldness and innovation, especially among early-career political scientists. Many important scholarly works initially defy easy contextualization and fit poorly into existing literatures, as the example of John Nash's paper shows. Tenure and promotion decisions that are based only or largely on citations garnered within five years of publication reward competence over lasting significance.

Research-Community Size

Scholars in larger research communities have an advantage over scholars in smaller fields of study: a larger pool of scholars who can cite them.[6] A paper cited in 16 of 100 articles published in a given year on the US Congress probably has less "impact" than a paper cited in eight of the 10 articles published that year on Pakistan, but it has double the count. Scholars producing incremental improvements in well-tilled fields thus tend to have larger citation counts than scholars producing novel insights in small but important or growing fields.

Gender

Maliniak, Powers, and Walter (2013) analyzed more than two decades of IR publications and found that—controlling for various factors such as publication venue, methodology, and tenure status—an article written by a woman receives 80% as many citations as a similar article written by a man. Women are less likely to be cited by the most influential articles and less likely to self-cite, possibly because they are penalized for self-promotion (Moss-Racusin and Rudman 2010).

Dion, Sumner, and Mitchell's (forthcoming) study shows that women are more likely than men to cite works by women in multiple social science journals. In general, men and mixed-author teams tend to under-cite women's work, although the size of the gender-citation gap varies according to the number of women in the field. As numbers of women scholars increase, work by women tends to be cited more by both men and women authors. They conclude that "citation practices are influenced by gender diversity" (Dion, Sumner, and Mitchell, forthcoming).

Biases in favor of men arise even if scholars genuinely cite those who influence them most. Colgan (2017) found that male IR instructors are less likely than female instructors to assign work by women scholars, whereas women also are less likely than men to assign their own work. Although there have been few

https://doi.org/10.1017/S104909651800094X Published online by Cambridge University Press

systematic studies in political science on whether other underrepresented scholars experience similar biases, it is highly plausible that such biases exist.

Coauthorship

Scholars who tend to coauthor can generate citations more easily than those who tend to solo author. Citation counts are not divided by the number of coauthors; that is, if a paper with five authors is cited once, each author receives one citation, not 0.2 citations. Higher citation counts for coauthoring scholars can exacerbate other biases. For example, Teele and Thelen (2017) demonstrated that all-male teams authored most of the collaborative work in 10 of the most prominent political science journals.

The disadvantages of GS do not foreclose its potential to serve as one indicator of scholarly impact, but we should understand the factors that affect decisions about whom to cite. If departments, deans, and granting institutions look at GS counts and external letter writers use them as a guide, then citation numbers effectively are being counted twice or more, thereby reducing the independence of external scholars' judgment. This double-counting is particularly problematic when the data-generating process that leads to higher and lower citation counts has not yet been fully understood. Moreover, because a GS profile is a public signal, it can have a disproportionate effect on opinions because a person seeing it knows that others also see it (Chwe 2016).

We recommend against the use of GS by external reviewers evaluating candidates for tenure and promotion. Their job is not to repeat information that anyone with access to the Internet can obtain. Rather, it is to do work that most department members, deans, and provosts—actors who will continue to use GS as a shortcut—cannot.

Strategic Citation

Scholars face an array of professional imperatives: they want their work to appeal to reviewers, be published, and garner citations. Thus, they may engage in several forms of strategic citation (Aizenman and Kletzer 2011).

Some books and articles are cited almost entirely for their flaws, not their importance. Authors often cite poorly executed studies, easy targets, and "straw-man" pieces to explain what they are arguing against (Nexon and Jackson 2015). Journalists and policy analysts, in particular, may cite scholarship that generates splashy headlines and overlook less provocative work of higher quality (Colgan 2016).

Citation choices also may be guided by expectations of likely peer reviewers. Strategic citations include fellow members of academic networks likely to be favorably predisposed, producing a bias against citing scholars who are critical or outside of those networks.

Moreover, strategic citations may be driven by calculations about which works editors and reviewers will expect to be listed in the bibliography. Journal articles are increasingly subject to fewer word limits. Omitting citations helps, but authors do not want to be criticized for missing key works. This fear encourages "drive-by citations"—that is, citing papers merely because similar papers cite those papers, regardless of their actual relevance. Anecdotally, scholars who are cited in this way often are those whose work is repeatedly cited erroneously for arguments they did not actually make.

Strategic citation leads scholars to cite authors of works already deemed important, not because of their relevance but rather because they are perceived as gatekeepers, hold key editorial positions, or reside in powerful departments. Junior scholars have expressed concerns that their submissions to journals will face rejection if they anger prominent scholars by criticizing or not mentioning their work. Like the biases discussed previously, strategic citation reproduces existing inequalities and disfavors underrepresented minorities, scholars from lower-ranked institutions, and those doing innovative work that does not fit neatly into existing literatures.

CONCLUSION

GS has advantages: it promotes consistency in research evaluation; encourages transparency, publicity, and openness; makes it easier to gain access to scholarly work; facilitates networking among scholars; and may provide incentives for quality over quantity. On the other hand, GS citation counts favor incremental work, scholars in larger research communities, male—and likely white—scholars, scholars who coauthor, and work that is cited strategically. Although it breaks down some doors, the uncritical use of GS entrenches long-existing inequalities in the political science discipline.

We recommend against the use of GS by external reviewers evaluating candidates for tenure and promotion.[7] Their job is not to repeat information that anyone with access to the Internet can obtain. Rather, it is to do work that most department members, deans, and provosts—actors who will continue to use GS as a shortcut—cannot. We rely on the judgment of external reviewers because they are experts in a candidate's field. They have a unique role and obligation to offer a reasoned, autonomous, and qualitative account of a scholar's contributions.

We are concerned that many academic departments and external reviewers—seduced by the ease of GS—are granting it unmerited importance in hiring, tenure, and promotion decisions. Because these decisions are vital for careers and lives, they should be based on data that are as accurate and as balanced as possible. We are not against using GS, but caution that it should be only one source of information used to evaluate scholars and only in conjunction with other means of assessment. As GS grows in importance, so does the need for autonomous scholarly judgment. ■

NOTES

1. See Reiter's (2016) report on the use of citation data for promotion decisions in departments with political scientists.
2. The authors of this article are not in complete agreement on all points.
3. See "About Google Scholar" available at https://scholar.google.com/intl/en/scholar/about.html (accessed April 5, 2018).
4. See www.hefce.ac.uk/rsrch/funding.

https://doi.org/10.1017/S104909651800094X Published online by Cambridge University Press

5. There are additional concerns about the accuracy of GS counts. GS appears to vacuum up virtually everything on the web without any quality control. Reiter (2016) suggested that this inclusivity leads to over-counting. For instance, Samuels (2011) found that one of his journal articles had 80 citations according to GS; however, six were duplicate entries and 52 were unpublished works. An outsider cannot improve the tool because Google does not disclose its algorithms. Over-counting can result from the existence of multiple online versions of an article, subtle variations in bibliographic format, and accidental inclusion of articles written by others. Scholars also can edit their own GS citation-count data, which may lead to manipulation of the citation count (Reiter 2016)—for instance, through adding or failing to delete false positives. GS also may under-count. Samuels (2011) also found that five of the 80 citations to his article in GS were books. However, according to the Google Books database—a separate database not linked to Google Scholar—the article was cited in 24 books. Under-counting of citations in books may decline as more material is entered into the GS database over time. (Samuels found no difference in book citations to the same article in 2017, whereas a difference did appear in 2010.) However, because the methodology is not public, self-correction over time cannot be assumed. To the extent that GS still under-counts citations that books receive or of articles in books, GS may understate the impact of scholars who publish more in books than in articles (Samuels 2013).
6. Reiter (2016) also pointed out variation in the size of the citation pool across political science subfields, rendering problematic comparisons of scholars in different subfields.
7. We know that many external reviewers will continue to use GS in the course of doing research about candidates for tenure and promotion. We are not recommending that they stop this practice but rather that they do not include GS and other citation counts in their written evaluations. However, the authors of this article are not in complete agreement on this point.

REFERENCES

Aizenman, Joshua, and Kenneth Kletzer. 2011. "The Life Cycle of Scholars and Papers in Economics—the 'Citation Death Tax.'" *Applied Economics* 43 (27): 4135–48.

Chwe, Michael Suk-Young. 2016. "Stereotypes Are More Powerful When People Like to Agree with Each Other." University of California, Los Angeles: Working paper. Last modified September 30. PDF file.

Colgan, Jeff. 2016. "Where Is International Relations Going? Evidence from Graduate Training." *International Studies Quarterly* 60 (3): 486–98.

———. 2017. "Gender Bias in International Relations Graduate Education? New Evidence from Syllabi." *PS: Political Science & Politics* 50 (2): 456–60.

Dion, Michelle, Jane Lawrence Sumner, and Sara McLaughlin Mitchell. Forthcoming. "Gendered Citation Patterns across Political Science and Social Science Methodology Fields." *Political Analysis*.

Ginther, Donna K., Walter T. Schaffer, Joshua Schnell, Beth Masimore, Faye Liu, Laurel L. Haak, and Raynard Kington. 2011. "Race, Ethnicity, and NIH Research Awards." *Science* 333 (6045): 1015–19.

Hendrix, Cullen. 2015. "Google Scholar Metrics and Scholarly Productivity in International Relations." *Duck of Minerva*, August 6. Available at http://duckofminerva.com/2015/08/google-scholar-metrics-and-scholarly-productivity-in-international-relations.html.

———. 2016. "To Be or Not to Be, 'Tis in Question." In "Scholarly Influence and the Shaping of International Relations Debates." *International Studies Quarterly Symposium*, March 31.

Jacsó, Peter. 2005. "Google Scholar: The Pros and the Cons." *Online Information Review* 29 (2): 208–14.

Kalyvas, Stathis N. 1999. "Wanton and Senseless? The Logic of Massacres in Algeria." *Rationality and Society* 11 (3): 243–85.

———. 2006. *The Logic of Violence in Civil War*. New York: Cambridge University Press.

Maliniak, Daniel, Ryan Powers, and Barbara F. Walter. 2013. "The Gender Citation Gap in International Relations." *International Organization* 67 (4): 889–922.

Milkman, Katherine L., Modupe Akinola, and Dolly Chugh. 2012. "Temporal Distance and Discrimination: An Audit Study in Academia." *Psychological Science* 23 (7): 710–17.

Moss-Racusin, Corinne A., John F. Dovidio, Victoria L. Brescoll, Mark J. Graham, and Jo Handelsman. 2012. "Science Faculty's Subtle Gender Biases Favor Male Students." *Proceedings of the National Academy of Sciences* 109 (41): 16474–9.

Moss-Racusin, Corinne A., and Laurie A. Rudman. 2010. "Disruptions in Women's Self-Promotion: The Backlash Avoidance Mode." *Psychology of Women Quarterly* 34 (2): 186–202.

Nexon, Daniel (ed.). 2016. "Scholarly Influence and the Shaping of International Relations Debates." *International Studies Quarterly Symposium*, March 3. Available at www.isanet.org/Publications/ISQ/Posts/ID/5107/Scholarly-Influence-and-the-Shaping-of-International-Relations-Debates.

Nexon, Daniel, and Patrick Thaddeus Jackson. 2015. "Academia Isn't Baseball." *Duck of Minerva*, August 10. Available at http://duckofminerva.com/2015/08/academia-isnt-baseball.html.

Ray, Debraj, and Arthur Robson. 2018. "Certified Random: A New Order for Coauthorship." *American Economic Review* 108 (2): 489–520.

Reiter, Dan. 2016. "Citation Count Data and Faculty Promotion." *Duck of Minerva*, February 18. Available at http://duckofminerva.com/2016/02/citation-count-data-and-faculty-promotion.html.

Samuels, David. 2011 "The Modal Number of Citations to Political Science Articles Is Greater Than Zero: Accounting for Citations in Articles and Books." *PS: Political Science & Politics* 44 (4): 783–92.

———. 2013. "Book Citations Count." *PS: Political Science & Politics* 46 (4): 785–90.

Stephan, Paula, Reinhilde Veugelers, and Jian Wang. 2017. "Reviewers Are Blinkered by Bibliometrics." *Nature* 544 (7651): 411–12.

Teele, Dawn Langan, and Kathleen Thelen. 2017. "Gender in the Journals: Publication Patterns in Political Science." *PS: Political Science & Politics* 50 (2): 433–47.

Wang, Dashun, Chaoming Song, and Albert-Laszlo Barabasi. 2013. "Quantifying Long-Term Scientific Impact." *Science* 342 (6154): 127–131.