



*Evidence Summary*

**Publons Peer Evaluation Metrics are not Reliable Measures of Quality or Impact**

**A Review of:**

Ortega, J. L. (2019). Exploratory analysis of Publons metrics and their relationship with bibliometric and altmetric impact. *Aslib Journal of Information Management*, 71(1), 124–136.  
<https://doi.org/10.1108/AJIM-06-2018-0153>

**Reviewed by:**

Scott Goldstein  
Web Librarian  
Appalachian State University Libraries  
Boone, North Carolina, United States of America  
Email: [goldsteinsl@appstate.edu](mailto:goldsteinsl@appstate.edu)

**Received:** 1 May 2019

**Accepted:** 14 July 2019

■ 2019 Goldstein. This is an Open Access article distributed under the terms of the Creative Commons-Attribution-Noncommercial-Share Alike License 4.0 International (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly attributed, not used for commercial purposes, and, if transformed, the resulting work is redistributed under the same or similar license to this one.

DOI: 10.18438/eblip29579

---

**Abstract**

**Objective** – To analyze the relationship between scholars' qualitative opinion of publications using Publons metrics and bibliometric and altmetric impact measures.

**Design** – Comparative, quantitative data set analysis.

**Setting** – Maximally exhaustive set of research articles retrievable from Publons.

**Subjects** – 45,819 articles retrieved from Publons in January 2018.

**Methods** – Author extracted article data from Publons and joined them (using the DOI) with data from three altmetric providers: Altmetric.com, PlumX, and Crossref Event Data. When providers gave discrepant results for the same metric, the maximum value was used. Publons data are described, and correlations are calculated between Publons metrics and altmetric and bibliometric indicators.

**Main Results** – In terms of coverage, Publons is biased in favour of life sciences and subject areas associated with health and medical sciences. Open access publishers are also over-represented. Articles reviewed in Publons

overwhelmingly have one or two pre-publication reviews and only one post-publication review. Furthermore, the metrics of significance and quality (rated on a 1 to 10 scale) are almost identically distributed, suggesting that users may not distinguish between them. Pearson correlations between Publons metrics and bibliometric and altmetric indicators are very weak and not significant.

**Conclusion** – The biases in Publons coverage with respect to discipline and publisher support earlier research and suggest that the willingness to publish one’s reviews differs according to research area. Publons metrics are problematic as research quality indicators. Most publications have only a single post-publication review, and the absence of any significant disparity between the scores of significance and quality suggest the constructs are being conflated when in fact they should be measuring different things. The correlation analysis indicates that peer evaluation in Publons is not a measure of a work’s quality and impact.

### Commentary

The study is the first in-depth examination of articles in Publons, a web platform that allows users to make public their peer-review reports for journals as well as rate articles on quality and relevance. Previous studies looking at Publons have focused on who uses the service rather than what articles are reviewed. This study sheds more light on two questions that have been asked in the literature. First, what is the relationship between traditional bibliometric indicators and peer subjective valuations (interpreted more broadly than refereed peer review as it encompasses other pre- and post-publication reviews)? In a comparable study focusing on Faculty of 1,000 (F1000) recommendations, Waltman and Costas (2014) found only a weak correlation between peer evaluations and citations, suggesting that the measures may capture different types of impact. Second, in what ways do the data providers used in scientometric analysis provide discrepant data? Zahedi and Costas (2018), for instance, conducted the most exhaustive comparison of

altmetric providers and found that coverage of publications between them varies widely.

This commentary relies on Perryman’s (2009) critical appraisal tool for bibliometric studies. The study fares very well when evaluated against it. For instance, the author clearly states the research questions after a literature review that motivates the need for a more in-depth look at Publons. The data are analyzed using appropriate statistical methods given the research objectives, and the results are graphically displayed in an appropriate manner given the types of analyses performed. The correlations between all the metrics are presented in a correlation matrix, which, although missing a colour legend, is an excellent way to visualize this type of data. Limitations of the data sources are also considered. Sufficient, though perhaps minimal, detail was provided on how the author scraped the data from Publons; it would be difficult for a novice researcher in this area to replicate the study. It is not exactly clear whether scraping the website violates Publons’ terms of service, which explicitly state that “The correct way to access these data is via our API.”

This study highlights a couple of implications for information professionals. Peer evaluations of scholarship outside of traditional (double-) blind peer review (pre- and post-publication review and open peer review as adopted by some journals during the refereeing process) is an emerging practice. Librarians should familiarize themselves with these new methods of review and be in a position to inform scholars of their opportunities and challenges (see Ford, 2016). Information professionals may also take away from this study a more measured view of the role peer review plays in the assessment of scholarly literature. Although this study suggests Publons metrics are not suitable quality indicators for reasons that are specific to the Publons platform, it would be wise to remember that peer review is not an infallible measure of quality, let alone impact. An evidence-based critical appraisal of a scholarly work does not simply consist of checking the peer review status.

## References

- Ford, E. (2016). Opening review in LIS journals: A status report. *Journal of Librarianship and Scholarly Communication*, 4(General Issue), eP2148.  
<https://doi.org/10.7710/2162-3309.2148>
- Perryman, C. (2009). *Evaluation tool for bibliometric studies*. Retrieved from <https://www.dropbox.com/l/scl/AAAL7LUZpLE90FxFnBv5HcnOZ0CtLh6RQrs>
- Waltman, L., & Costas, R. (2014). F1000 recommendations as a potential new data source for research evaluation: A comparison with citations. *Journal of the Association for Information Science and Technology*, 65(3), 433–445.  
<https://doi.org/10.1002/asi.23040>
- Zahedi, Z., & Costas, R. (2018). General discussion of data quality challenges in social media metrics: Extensive comparison of four major altmetric data aggregators. *PLoS ONE*, 13(5), e0197326.  
<https://doi.org/10.1371/journal.pone.0197326>