

To what extent has open access (OA) publishing benefited researchers in lower-income countries? A global analysis of reference patterns 1980-2020

Henrik Karlstrøm, Fredrik N. Piro and Dag W. Aksnes

henrik.karlstrom@nifu.no; fredrik.piro@nifu.no; dag.w.aksnes@nifu.no

0000-0002-9942-4326, 0000-0002-8108-778X, 0000-0002-1519-195X

NIFU – Nordic Institute for studies in Innovation, Research and Education, Norway

Making scientific literature freely available to everyone is a main objective of the open access (OA) movement. This may be especially crucial for researchers in lower-income countries who often face barriers due to high subscription costs. In this paper, we address this issue by analyzing the reference lists of scientific publications worldwide over time. Our study focuses on key issues, including whether researchers from lower-income countries refer to fewer previous publications when they publish and how this trend evolves over time. We also investigate whether researchers from lower-income countries rely more on literature that is openly available through different OA routes compared to other researchers. Our study reveals that the proportion of references that are OA is increasing over time for all publications and country groups. However, publications from lower-income countries have a higher growth rate of OA-references, suggesting that the emergence of OA-publishing has been particularly advantageous for researchers in these countries.

1. Introduction¹

A fundamental objective of the open access (OA) movement has been to make scientific literature freely available to everyone. Recent studies indicate that the percentage of OA publications indexed in Web of Science (WoS) is approaching 50%; in the period 2015-2019 42.9% of WoS publications were OA (Simard et al., 2022).

In the debate about the benefits of OA, lower-income countries have typically been chosen as a case to illustrate the advantages (and disadvantages) of OA. Changing the costs of publishing from subscription fees to author payments (Budzinski et al., 2020), has inevitably sparked a discussion about whether lower-income countries, as producers of research, would be disadvantaged by no longer being affordable to publish in journals (Nature, 2022). Our study is situated in the potential benefits of OA from the consumer perspective, i.e., in the use of knowledge. From this perspective it is reasonable to foresee that the greatest use, and hence benefit, would be in countries where access to scientific publications initially was poor. Several recent studies have explored country and regional differences in OA publishing, with a low-income perspective in mind, both from the producing side, finding that lower-income countries have the highest shares of OA publishing (e.g., Iyandemye & Thomas, 2019), and from the consuming side, the highest share of OA citations (Simard et al., 2022).

Our study is grounded in the latter perspective. In addition to presenting descriptive findings of OA reference patterns, we also analyze how and to what extent OA is beneficial to lower-income countries. Here, we explore whether the potential increase in use of OA publications and references in lower-income countries, relative to high-income countries, covary with other

¹ An extended version of this paper has been posted on SocArXiv Papers: <https://doi.org/10.31235/osf.io/ecgzh>

indicators that express evidence of better research capacity or greater inclination to be updated on the frontiers of science.

The *raison d'être* for the expectation that OA is more beneficial to research systems in lower-income countries has to do with these countries' initial low investments in R&D. More specifically, in several developing countries, foremost in Africa, the access to scientific literature has been limited (Abrahams et al., 2008; Chan & Costa, 2005). One main problem with low R&D investments, is the low ability to cover costs for (very expensive) journal subscriptions (Aronson, 2004; Dulle & Minishi-Majanja, 2009; Matheka et al., 2014). Prior to the OA wave, Chan & Costa (2005) illustrate the problem with data from a WHO survey, showing that in countries with a GNP per capita of less than USD 1000, more than 50% of academic institutions did not have *any* journals subscriptions, while in countries with a GNP per capita in the range USD 1000-3000, 34% of academic institutions did not have any subscriptions (while 34% reported to have an average of two subscriptions).

The rationale for studying the number of references is based on the simple logic that accessibility to scientific literature is a determinant to how many relevant papers an author/group of authors can meaningfully enter to the paper. In high-income countries, the access was already in place before OA, whilst in lower-income countries (foremost African countries) this has often not been the case, where there has historically been poor access to internet and computer facilities, not enabling the research to fully utilize the increasingly available online literature (Adcock & Fottrell, 2008; Chan & Costa, 2005). One may therefore reason that with the increasing numbers of OA papers available, the numbers of references given in a paper may have increased relatively more in lower-income countries than in higher-income countries.

2. Data and methods

The study is based on Web of Science (WoS) data. The WoS Core collection, which we use in this study, now covers Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index, Emerging Sources Citation Index, Conference Proceedings citation indexes, and Book Citation Index. We have applied a local version of WoS maintained by the Norwegian Agency for Shared Services in Education and Research.

The analyses have been carried out at a global level. Because transnational collaborations make it impossible to identify the country source of a given reference in a publication, the study is limited to regular articles with only one single author country, in total 32.7 mill. articles during the period 1980-2020 and 1.29 billion references in the reference lists of these articles. The corresponding global total, also including internationally co-authored articles is 42.0 mill. Thus, the analysis covers the large majority (78 per cent) of the WoS indexed articles published during the period. This is in accordance with a recent study showing that in 2020, 25 per cent of the world-production of articles involved international collaboration (Aksnes & Sivertsen 2022). We are classifying author countries into four income groups ("high", "upper middle", "lower middle" and "low" income), based on the World Bank's classification of countries by 2020 gross national income (GNI) per capita (World Bank, 2022).

The number of papers included in the analysis has increased from approx. 380,000 in 1980 to 1,702,000 in 2020. The number of papers from the high-income group is by far the highest with 24,020,000 articles over the entire period. The corresponding figures for the other groups are: Upper middle-income: 6,554,000 articles; lower middle-income 2,094, 000; low-income 2941.

As can be seen, there are remarkably few publications from the low-income countries. This is both because these countries have very limited national budgets for research activities and because most of the research is carried out in collaboration with researchers from other countries, and results in internationally coauthored publications (Aksnes & Sivertsen, 2022), which are not included in the present study.

Data on OA status is retrieved from the Unpaywall database. Unpaywall operates with several flavors of OA. Since we are concerned only with the possibility of reading the contents of a publications for use as a reference in a publication and not the access license indicated by the different types of OA, we considered all forms of OA as one category for the purposes of this study.

Not all publications present in the reference lists of the publication set could be assigned an OA status. This is due to several factors: 1) not all references in a publication are to scientific publications, 2) WoS only indexes a subset of all scientific publications, 3) not all references have a DOI making matching with Unpaywall possible, 4) not all scientific publications have an OA status in Unpaywall.

Having collected all publications in WoS that fit the criteria outlined above (published 1980-2020, regular articles, single country authorship, having a DOI), the publications were matched against the Unpaywall database to obtain their OA status.

Analyses have been carried out at field levels, but the results have not been included in this short paper.

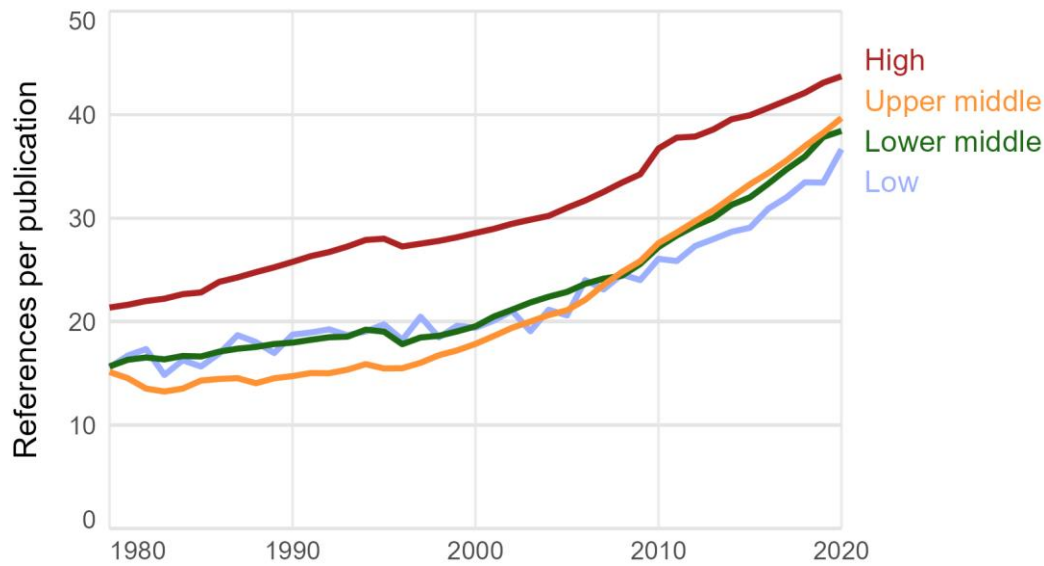
3. Results

3.1. Overall reference patterns over time

The number of references per article has increased considerably during the period 1980-2020. In 1980 an average article contained 19.8 references compared with 43.2 in 2020. This general upward trend is present in all fields. At the same time, there are large differences across fields, where for example biology has much longer reference lists on average than mathematics and statistics (to be shown in the final paper).

Figure 1 shows that articles from high-income countries have many more references on average than publications from other income groups. This holds over the entire 40-year period, but the gap has been reduced in recent years. Interestingly, there are relatively small differences across the groups of low- and middle-income countries.

Figure 1: Average number of references per publication, 1980-2020, by country income group.



The general trends from Figure 1 also hold when the issue is analyzed at field levels: across all fields and income groups the number of references per paper is increasing steadily and there is a gap between the high-income countries and the middle- and low-income countries. However, some field-specific differences exist (to be shown in the final paper).

3.2. Changes in OA and non-OA reference pattern over time

Within the period under consideration another trend has happened concurrently with the increase in the number of references per paper – the growth of openly available scientific publications. In 2005, publications from high-income countries had on average 7.5 OA references and 24.2 non-OA references (Table 1). These numbers increased to 16.5 OA and 30.3 non-OA references in 2020. Thus, the growth is stronger for OA than for non-OA references in absolute as well as relative terms. A similar pattern is shown for middle- and low-income countries, which have been combined in the table.

Table 1. Average number of references per publication, 2005-2020, by country income group and access type.

	High-income		Middle/low-income	
	Non-OA	OA	Non-OA	OA
2005	24.2	7.5	18.6	7.4
2006	24.7	7.7	20.0	8.5
2007	25.5	7.9	20.4	7.6
2008	26.5	8.2	21.0	8.2
2009	27.0	8.7	20.8	7.9
2010	28.6	9.3	22.0	8.1
2011	29.3	9.7	22.5	8.2
2012	29.5	10.0	23.3	9.2
2013	29.7	10.4	23.5	8.9
2014	30.2	10.9	24.3	9.5
2015	30.2	11.4	24.3	10.3
2016	30.2	12.4	25.2	10.7
2017	30.5	13.3	25.7	11.6
2018	30.4	14.1	26.2	12.5
2019	30.5	15.4	26.3	13.2
2020	30.3	16.5	26.9	14.9

Table 2 demonstrates that the compound average growth rate (CAGR) is positive for both types of references, but highest for OA references for all income groups, and highest for the two lowest income groups.

The increased growth rate of OA references, particularly among publications from lower-income countries, indicates a convergence of reference practices between researchers from different income groups.

Table 2. Compound Average Growth Rate by income group and access type of references, 2005-2020.

Income group	Non-OA	OA
Low-income	35 %	42 %
Lower middle-income	28 %	33 %
Upper middle-income	26 %	32 %
High-income	13 %	18 %

The calculations have also been carried out at field levels. This analysis reveals that the overall OA patterns identified for groups of countries are present also at field levels, but in some fields, the patterns do not follow the general tendencies for country groups, making the conclusions less generic. We have also explored whether increased use of OA publications is also followed by improved access to scientific literature, being expressed in more recent publications being cited in the reference lists. Due to lack of space, these results have not been included in the present version.

4. Discussion & conclusions

This study has shown that the reference practice of scientists has undergone major changes the last 40 years. The length of the reference list has more than doubled. Publications from low-income countries used to include much fewer references than publications from high-income countries. This gap is reduced during the period.

Another major change relates to the transformation of scientific communication where publications are much more easily accessible. Here, OA is an important part of the development. This study has revealed that the increased availability of literature through OA publishing has affected researchers from low- and high-income countries differently. The growth rate of references which are OA is strongest for low-income countries. This suggests that the rise of OA publishing has been particularly beneficial for researchers in low-income countries. In this way a fundamental objective as well as justification of the OA movement has been accomplished.

Still, we observe that even today the large majority of the references of the low-income countries go to closed literature. One may ask why this is the case and why the OA differences across country groups are not even larger. Two factors are relevant to consider in this context.

First, researchers in lower-income countries are facing the same situation as researchers in high-income countries when they are collecting literature. Only parts of the relevant previous literature on a particular topic is openly available. Thus, based on such considerations one would not expect radically different patterns.

Second, researchers in lower-income countries may also have access to the closed literature. Of course, there are universities and libraries in the lower-income countries with subscription based access to the literature. Second, what appears as closed literature in this study, may still be openly available. In particular, it should be noted that articles available through academic social networks only (e.g. ResearchGate and Academia.edu) are not counted as OA, due to the criteria applied by Unpaywall (Piwowar et al., 2018). Full texts of a very large body of literature are available through these websites (Taylor, 2022). Another option is to contact the authors directly to get printed version of articles or pdfs sent by e-mail. Finally, it is known that not all references cited in publications have been read by the authors citing them, and reference copying is widespread (Simkin, & Roychowdhury, 2007).

As mentioned in the introduction, many previous studies have been conducted on the role of OA from a global equity perspective. However, most studies have been limited to the producing side, while research on the consumer side is limited (i.e., the use of literature as reflected in reference patterns).

One exception is the study by Simard et al. (2022), which showed that low-income countries have a higher uptake of OA than high-income countries. Although it has similarities to our study, it also has fundamental differences in the methodological approach. Simard et al. assigned papers to country groups based on the address affiliation of the first and corresponding authors only. Thus, publications with coauthors from additional countries were also included. This may affect the findings for low-income countries, as their scientific production is limited, and international collaboration is extremely widespread. Such a methodological approach may be problematic, and our study is therefore limited to articles that do not involve international collaboration. Nevertheless, the overall findings of our study

resemble those of Simard et al., in that the methodological disparities did not lead to a different main conclusion.

While open access (OA) has been particularly beneficial for researchers in lower-income countries by providing unrestricted access to publications, a new problem has emerged due to the shifting of publishing costs from subscription fees to author payments (Budzinski et al., 2020). This means that researchers in lower-income countries, when being the producers of research, will be disadvantaged as they may no longer be able to afford the fee for publishing (Nature, 2022).

Open science practices

The data analyzed in this study have been retrieved from the Web of Science database, licensed by the Norwegian Agency for Shared Services in Education and Research from Clarivate. Investigations will be carried out whether the licensing agreement permits distribution of publication metadata from Web of Science in aggregated form.

Competing interests

The authors have no competing interests.

References

Abrahams, L., Burke, M., Gray, E. & Rens, A. (2008). Opening Access to Knowledge in Southern African Universities. Southern African Regional Universities Association (SARUA): Study Series 2008. ISBN: 978-0-9814099-8-6.

Abrahams, L., Burke, M., Gray, E. & Rens, A. (2008). *Opening Access to Knowledge in Southern African Universities*. Southern African Regional Universities Association (SARUA): Study Series 2008. ISBN: 978-0-9814099-8-6.

Adcock, J., & Fottrell, E. (2008). *The North-South information highway: case studies of publication access among health researchers in resource-poor countries*. Global Health Action, 1 (1). <https://doi.org/10.3402/gha.v1i0.1865>

Aksnes, D.W. & Sivertsen, G. (2022). Patterns and trends in country level research collaboration over four decades. In N. Robinson-Garcia, D. Torres-Salinas, & W. Arroyo-Machado (Eds.), *26th International Conference on Science and Technology Indicators*, STI 2022(sti2275). <https://doi.org/10.5281/zenodo.6974640>

Aronson, B. (2003). Improving Online Access to Medical Information to Low-Income Countries. *New England Journal of Medicine*, 350(10), 966-968.

Budzinski, O., Grebel, T., Wolling, J. & Zhang, X. (2020). Drivers of article processing charges in open access. *Scientometrics*, 124, 2185–2206.

Chan, L., & Costa, S. (2005). *Participation in the global knowledge commons. Challenges and opportunities for research dissemination in developing countries*. New Library World, 106(3-4), 141-163.

Dulle, F.W. & Minishi-Majanja, M.M. (2009). Researchers' perspectives on open access scholarly communication in Tanzanian public universities. *South African Journal of Information Management*, 11(4), a413.

Iyandemye, J. & Thomas, M.P. (2019). Low income countries have the highest percentages of open access publication: A systematic computational analysis of the biomedical literature. *PLoS ONE*, 14(7), e0220229.

Matheka DM, Nderitu J, Mutonga D et al (2014). Open access: academic publishing and its implications for knowledge equity in Kenya. *Globalization and Health*. Dec;10(1):1-5.

Nature (2022). Open-access publishing fees deter researchers in the global south. 22 February, 2022. <https://www.nature.com/articles/d41586-022-00342-w>

Piwowar, H., Priem, J., Lariviere, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., West, J., & Haustein, S. (2018). The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. *Peerj*, 6, Article e4375. <https://doi.org/10.7717/peerj.4375>

Simard M-A, Ghiasi G, Mongeon P & Larivière V. (2022). National differences in dissemination and use of open access literature. *PLoS ONE*, 17(8), e0272730. <https://doi.org/10.1371/journal.pone.0272730>

Simkin, M. V., & Roychowdhury, V. P. (2007). A mathematical theory of citing. *Journal of the American Society for Information Science and Technology*, 58(11), 1661-1673. <https://doi.org/10.1002/asi.20653>

Taylor, J. (2022). The Shady Margins of Open Access: Gray OA and Academic Social Network Sites. *Technical Services Quarterly*, 39(4), 428-439. <https://doi.org/10.1080/07317131.2022.2125679>

World Bank (2022). How does the World Bank classify countries? Available at: <https://datahelpdesk.worldbank.org/knowledgebase/articles/378834-how-does-the-world-bank-classify-countries>