

# The methodology used for the Times Higher Education World University Rankings' citations metric can distort benchmarking

*The Times Higher Education World University Rankings can influence an institution's reputation and even its future revenues. However, Avtar Natt argues that the methodology used to calculate its citation metrics can have the effect of distorting benchmarking exercises. The fractional counting approach applied to only a select number of papers with high author numbers has led to a situation whereby the methodologists have unintentionally discriminated against certain types of big science paper. This raises questions about the benchmarking and also reiterates the importance of such rankings maintaining transparency in their data and methods.*

Since the [2017-2018 results of the Times Higher Education World University Rankings](#) (THEWUR) were released last month, academic policymakers will inevitably have been asking questions about the position of their own institutions. After closer examination of the [methodology](#) of this report (the same as used for the [2016-17 THEWUR](#)) a concern may be raised over the citations metric that accounts for 30% of an institution's overall score. Given these rankings can influence an institution's reputation and future revenues, there is hope that raising this concern may lead to greater openness about the methods and data produced.

In the [methodology of THEWUR's 2015-16 report](#), papers with more than 1,000 authors were excluded from the citations metric calculation. The total number of papers excluded was small (649) but when the methodology refers to such papers as "having a disproportionate impact on the citation scores of a small number of universities", this does invite curiosity.

Nevertheless, for the 2016-17 report the methodologists, in conjunction with Elsevier, opted to include those papers with more than 1,000 authors but with an important amendment. In their words:

"[We] have developed a new fractional counting approach that ensures that all universities where academics are authors of these papers will receive at least 5 per cent of the value of the paper, and where those that provide the most contributors to the paper receive a proportionately larger contribution."

*([Times Higher Education](#))*

Fractional counting is not the concern here (the [Leiden rankings](#) use it) but rather that fractional counting has only been applied to a miniscule number of papers. For each of the last two reports, we are estimating somewhere in the region of 725 select papers out of 12 million published outputs when the data is collected for the citations metric. A logical conclusion to arrive at is that those papers with an enormous number of authors and an enormous number of citations can distort benchmarking. Whether intended or not, Elsevier/THEWUR have classified a new type of paper: the mega-paper that has at least 1,000 authors and that, within a defined (yet limited) time period, receives a number of citations so enormous that it can distort citation benchmarking exercises.

To put this argument to the test I identified a sample of papers used for the 2016-17 citations ranking. Using Scopus classifications, outputs published 2011-2015 were retrieved with a focus on journal articles from UK institutions only. I accept that my retrieval date, 10 June 2017, was not the same as that of the methodologists but the retrieval strategy did still serve its purpose. This was done by looking at different

types of mega-authorship (papers with at least 100 authors or at least 1,000 authors) and how they combined with papers awarded at least 100 citations or at least 1,000 citations. The spread was as follows:

Year	2011	2012	2013	2014	2015	Total
Outputs	137,789	142,513	147,762	150,648	153,562	732,274
Outputs $\geq$ 100 citations	3,834	2,953	1,847	1,042	380	10,056
Outputs $\geq$ 100 citations and $\geq$ 100 authors	81	81	69	49	35	315
Outputs $\geq$ 100 citations and $\geq$ 1,000 authors	10	16	11	3	2	42
Outputs $\geq$ 1,000 citations	47	46	22	7	4	126
Outputs $\geq$ 1,000 citations and $\geq$ 100 authors	2	13	0	4	2	21
Outputs $\geq$ 1,000 citations and $\geq$ 1,000 authors	0	3	0	0	0	3

**Table 1: Scopus data based on 2011-2015 outputs from UK affiliated institutions**

Table 1 reveals that out of the sample retrieved, fractional counting made its biggest difference for 42 papers, each receiving at least 100 citations and each with at least 1,000 authors. Much of the readership will not be surprised to hear that 40 of the 42 relevant papers were classified by Scopus as belonging to the subject area of Physics and Astronomy. More interesting was that all of these 40 papers were CERN-related (the European Organization for Nuclear Research) and involved the group of authors from the Atlas Collaboration or CMS Collaboration. By way of comparison, mega-authorship papers escaping fractional counting yet receiving an enormous number of citations were scrutinised. This was done by retrieving the papers with at least 1,000 citations and at least 100 authors. From this new set of 21 papers, 14 received the Scopus subject area classification of Medicine, with six affiliated to the Bill and Melinda Gates-funded Global Burden of Disease Study.

“So what?” you might think. Well, here’s where the methodology for the citations metric rears its head again. An evident commonality (among papers of 1,000 authors or more) was the presence of Russell Group institutions or UK institutions with a reputation for research intensity. When looking at their data, such institutions’ citations metrics appear able to withstand the intervention of the methodologists due to their higher number and wider spread of overall citations. The presumption was subsequently made that UK institutions with a more modest research profile could be subject to volatility in their citation metric. The cases of two UK higher education institutions are particularly noteworthy. Firstly, one particular CERN-related paper received 3,391 citations and had 2,891 authors, six of whom were affiliated to the University of the West of England. Based on citation counts, this particular paper was highly valuable for the University of the West of England as it was worth approximately 13% of its total citations during the aforementioned period of study. However, in this extreme case, THEWUR’s methodology meant that fractional counting played a part in the University of the West of England not going beyond a citation metric of 38.3 in 2017 and 32.6 in 2018.

The second extreme example concerns Anglia Ruskin University and its one author who appeared in four of

the six Global Burden of Disease Study papers mentioned above. Based on the data, these four papers averaged 411 authors per paper and 2,310 citations per paper. When the four papers and their citation counts were considered alongside Anglia Ruskin's other outputs, they were found to be worth 45% of the university's total citations. The resulting effect on the THEWUR citations metric was a 2017 score of 99.2 and a 2018 score of 99.4. The combination of highly cited papers escaping fractional counting at an institution susceptible to volatility in its metrics served to dramatically improve its ranking. Standardised or not, an extreme outlier is an extreme outlier.

A norm of citation metrics is that the highest cited papers are rewarded rather than being disregarded as outliers. Of course, one can also sympathise with the argument that institutions collaborating in modern forms of big science should enjoy the spoils that come with it. Yet a big issue for institutional rankings based on citation data is how to treat papers with enormous citation counts and enormous (as well as confusing) numbers of institutional affiliations. Dividing citations based on the number of times an institution appears in the author list certainly has a democratic appeal. But it also raises a different set of problems because it is dependent on the citation database used and its particular strengths and weaknesses. While the examples I provide may be on the extreme side, there is still curiosity in how the institutional citation metrics would have looked if the fractional counting tweak was not applied. With the best of intentions, the methodologists resolved one issue but created another. One can observe the different calculations that were applied to relevant papers from Physics and Astronomy compared to those of Medicine, for example. A conclusion can thus be drawn that the methodologists have unintentionally discriminated against certain types of big science paper.

For my part, it was the ramifications of treating such a small number of papers differently for benchmarking that motivated this post. If the data doesn't come out right, should outlier papers be treated differently? Further, should the impact of such changes matter less because of the institution the data impacts? Exercises like the THEWUR are not going anywhere and if they are to remain so influential there should be appropriate peer scrutiny. This includes sharing the methods and data produced with more than just auditors. Scopus should at least let its subscribers download the citation data used for benchmarking and THEWUR should display further commitment to transparency in its data and methods, rather than focusing on the production of glossy reports that so dazzle policymakers.

*Note: This article gives the views of the author, and not the position of the LSE Impact Blog, nor of the London School of Economics. Please review our [comments policy](#) if you have any concerns on posting a comment below.*

*Featured image credit: [calculator](#) by Anssi Koskinen (licensed under a [CC BY 2.0](#) license). **About the author Avtar Natt** is a Subject Librarian at Goldsmiths. He wrote the majority of this post based at the University of Bedfordshire as an Academic Liaison Librarian. His interests overlap his background in sociology and information science.*

