

LSE statement on the responsible use of citation metrics 2016

The LSE adopts the following principles on the use of citation metrics to assess research publications:

The Leiden Manifesto for Research Metrics (2015)¹

1. Quantitative evaluation should support qualitative, expert assessment. Quantitative metrics can challenge bias tendencies in peer review and facilitate deliberation. This should strengthen peer review, because making judgements about colleagues is difficult without a range of relevant information. However, assessors must not be tempted to cede decision-making to the numbers. Indicators must not substitute for informed judgement. Everyone retains responsibility for their assessments.
2. Measure performance against the research missions of the institution, group or researcher. Programme goals should be stated at the start, and the indicators used to evaluate performance should relate clearly to those goals. The choice of indicators, and the ways in which they are used, should take into account the wider socio-economic and cultural contexts. Scientists have diverse research missions. Research that advances the frontiers of academic knowledge differs from research that is focused on delivering solutions to societal problems. Review may be based on merits relevant to policy, industry or the public rather than on academic ideas of excellence. No single evaluation model applies to all contexts.
3. Protect excellence in locally relevant research. In many parts of the world, research excellence is equated with English-language publication. Spanish law, for example, states the desirability of Spanish scholars publishing in high-impact journals. The impact factor is calculated for journals indexed in the US-based and still mostly English-language Web of Science. These biases are particularly problematic in the social sciences and humanities, in which research is more regionally and nationally engaged. Many other fields have a national or regional dimension – for instance, HIV epidemiology in sub-Saharan Africa.

This pluralism and societal relevance tends to be suppressed to create papers of interest to the

¹ <http://www.leidenmanifesto.org/>

gatekeepers of high impact: English-language journals. The Spanish sociologists that are highly cited in the Web of Science have worked on abstract models or study US data. Lost is the specificity of sociologists in high-impact Spanish-language papers: topics such as local labour law, family health care for the elderly or immigrant employment⁵. Metrics built on high-quality non-English literature would serve to identify and reward excellence in locally relevant research.

4. Keep data collection and analytical processes open, transparent and simple. The construction of the databases required for evaluation should follow clearly stated rules, set before the research has been completed. This was common practice among the academic and commercial groups that built bibliometric evaluation methodology over several decades. Those groups referenced protocols published in the peer-reviewed literature. This transparency enabled scrutiny. For example, in 2010, public debate on the technical properties of an important indicator used by one of our groups (the Centre for Science and Technology Studies at Leiden University in the Netherlands) led to a revision in the calculation of this indicator⁶. Recent commercial entrants should be held to the same standards; no one should accept a black-box evaluation machine.

Simplicity is a virtue in an indicator because it enhances transparency. But simplistic metrics can distort the record (see [principle 7](#)). Evaluators must strive for balance — simple indicators true to the complexity of the research process.

5. Allow those evaluated to verify data and analysis. To ensure data quality, all researchers included in bibliometric studies should be able to check that their outputs have been correctly identified. Everyone directing and managing evaluation processes should assure data accuracy, through self-verification or third-party audit. Universities could implement this in their research information systems and it should be a guiding principle in the selection of providers of these systems. Accurate, high-quality data take time and money to collate and process. Budget for it.

6. Account for variation by field in publication and citation practices. Best practice is to select a suite of possible indicators and allow fields to choose among them. A few years ago, a European group of historians received a relatively low rating in a national peer-review assessment because they wrote books rather than articles in journals indexed by the Web of Science. The historians had the misfortune to be part of a psychology department. Historians and social scientists require books and national-language literature to be included in their publication counts; computer scientists require conference papers be counted.

Citation rates vary by field: top-ranked journals in mathematics have impact factors of around 3; top-ranked journals in cell biology have impact factors of about 30. Normalized indicators are required, and the most robust normalization method is based on percentiles: each paper is weighted on the basis of the percentile to which it belongs in the citation distribution of its field (the top 1%, 10% or 20%, for example). A single highly cited publication slightly improves the position of a university in a ranking that is based on percentile indicators, but may propel the university from the middle to the top of a ranking built on citation averages⁷.

7. Base assessment of individual researchers on a qualitative judgement of their portfolio. The older you are, the higher your *h*-index, even in the absence of new papers. The *h*-index varies by field: life scientists top out at 200; physicists at 100 and social scientists at 20–30. It is database dependent: there are researchers in computer science who have an *h*-index of around 10 in the Web of Science but of 20–30 in Google Scholar⁹. Reading and judging a researcher's work is much more appropriate than relying on one number. Even when comparing large numbers of researchers, an approach that considers more information about an individual's expertise, experience, activities and influence is best.

8. Avoid misplaced concreteness and false precision. Science and technology indicators are prone to conceptual ambiguity and uncertainty and require strong assumptions that are not universally

accepted. The meaning of citation counts, for example, has long been debated. Thus, best practice uses multiple indicators to provide a more robust and pluralistic picture. If uncertainty and error can be quantified, for instance using error bars, this information should accompany published indicator values. If this is not possible, indicator producers should at least avoid false precision. For example, the journal impact factor is published to three decimal places to avoid ties. However, given the conceptual ambiguity and random variability of citation counts, it makes no sense to distinguish between journals on the basis of very small impact factor differences. Avoid false precision: only one decimal is warranted.

9. Recognize the systemic effects of assessment and indicators. Indicators change the system through the incentives they establish. These effects should be anticipated. This means that a suite of indicators is always preferable — a single one will invite gaming and goal displacement (in which the measurement becomes the goal). For example, in the 1990s, Australia funded university research using a formula based largely on the number of papers published by an institute. Universities could calculate the 'value' of a paper in a refereed journal; in 2000, it was Aus\$800 (around US\$480 in 2000) in research funding. Predictably, the number of papers published by Australian researchers went up, but they were in less-cited journals, suggesting that article quality fell.

10. Scrutinize indicators regularly and update them. Research missions and the goals of assessment shift and the research system itself co-evolves. Once-useful metrics become inadequate; new ones emerge. Indicator systems have to be reviewed and perhaps modified. Realizing the effects of its simplistic formula, Australia in 2010 introduced its more complex Excellence in Research for Australia initiative, which emphasizes quality.

It further adopts the two following principles from San Francisco Declaration on Research Assessment (2012).

11. Be explicit about the criteria used to reach hiring, tenure, and promotion decisions, clearly highlighting, especially for early-stage investigators, that the scientific content of a paper is much more important than publication metrics or the identity of the journal in which it was published.

12. For the purposes of research assessment, consider the value and impact of all research outputs (including datasets and software) in addition to research publications, and consider a broad range of impact measures including qualitative indicators of research impact, such as influence on policy and practice. [Note that given the wide range of indicators that can be used to assess research impact that it is not possible to prescribe a comprehensive set of indicators].

Review schedule

Review interval	Next review due by	Next review start
Every 1 year	31/12/2021	01/08/2021

Version history

Version	Date	Approved by	Notes
1	25 May 2016	Academic Board	

Links

Reference	Link

Contacts

Position	Name	Email	Notes
Research Support Manager	Helen Porter	h.porter@lse.ac.uk	
Head of Research Policy	Jo Hemmings	j.hemmings@lse.ac.uk	

Communications and Training

Will this document be publicised through Internal Communications?	Yes/ No
Will training needs arise from this policy	Yes/ No
If Yes, please give details	