University of Wolverhampton Research Indicators (Metrics) Policy

The aim of this document is to ensure the responsible use of impact indicators (metrics), when relevant. With this policy the University demonstrates its support for DORA (San Francisco Declaration on Research Assessment) to which the University became a signatory in 2020.

The University of Wolverhampton will avoid any implication that citation-based indicators or alternatives "measure" the quality of research. It will seek to use the term "Indicator" in preference to "metric" or "measure" as part of this. This reflects that indicators can give indirect information about likely scholarly or other impacts but never directly measure them. The University of Wolverhampton fully endorses the Metric Tide report guidelines for dimensions of metrics that should be considered.

- **Robustness**: basing metrics on the best possible data in terms of accuracy and scope
- **Humility**: recognising that quantitative evaluation should support – but not supplant – qualitative, expert assessment
- **Transparency**: keeping data collection and analytical processes open and transparent, so that those being evaluated can test and verify the results
- **Diversity**: accounting for variation by field, and using a range of indicators to reflect and support a plurality of research and researcher career paths across the system
- **Reflexivity**: recognising and anticipating the systemic and potential effects of indicators, and updating them in response.

The University of Wolverhampton's mission includes research and teaching as well as scholarship contributing to regional economic, health, social and cultural development. This document applies primarily to those pursuing research. Scholarly impact indicators are not relevant to academics that focus on teaching and regional development. They also have little relevance to those researching topics that legitimately have primary impact and interest within the local community.

The University of Wolverhampton will always permit, but never require, those being evaluated to present indicators in support of any claims for the quality or impact of their work. Recognising that academic work can have long term or hidden impacts, the absence of high indicator scores of any type will never be used by managers as evidence that work has had little impact. Academics are encouraged to produce the highest quality and most impactful work possible, and all indicators considerations are secondary to this. Indicators should always support a narrative impact claim and never replace it.

Recruitment

The University of Wolverhampton recognises that many academics work in specialist areas that no Wolverhampton employees would have the expertise to fully assess. This is particularly critical during recruitment, when decision makers may not have insufficient expertise or time to read and effectively evaluate the works of all applicants. In addition to seeking external input through references, the University will encourage applicants to explain their publishing or creative output strategy (e.g., artworks, performances) as part of their applications and make a claim for the value or impact of their work. Applicants may, if they wish, provide quantitative or other evidence in support of their narrative claim for the value of their work, such as citation counts, the prestige of the publishing journal or scholarly press (books), or published book reviews. They may also wish to present career citation indicators as evidence for the overall value or impact of their work. Whilst the support of indicators may strengthen an applicant's impact claim, their absence will not be taken as evidence that their work has had no impact.

Promotions

The rules for recruitment also apply to promotions. University often solicits the opinions of external experts as part of its promotions process, some of whom may include indicators as part of their evaluations. These indicators will be ignored unless they are presented as supporting evidence for a specific claim. If used, they will be re-evaluated in the context of the advice in this document, paying particular attention to diversity, age and field difference issues.

Self-evaluation

Research-active academics at the University of Wolverhampton are encouraged but not required, for their own self-evaluation purposes, to annually monitor citation and attention indicators for their work, if relevant in their field. This may help them to detect publishing topics or strategies that find a receptive audience to pursue in the future.

Publication venues

Academics at the University of Wolverhampton are encouraged to publish their work in the most appropriate venues, paying attention to the size and nature of the audience that each venue will attract. This includes journals and book publishers, as well as art galleries and performance venues. Publishing in prestigious venues, such as high reputation journals or publishers, is encouraged to attract rigorous peer review and a large appropriate audience. Nevertheless, valid reasons for choosing alternative outlets are welcomed. Publishing in predatory journals or conferences that lack effective peer review is valueless and is strongly discouraged.

Academics that write journal articles may claim that their work is published in a relevant prestigious journal as part of their evidence about the article's value. The use of Journal Impact Factors (JIFs) is discouraged because they vary over time, are not calculated robustly, and are greatly affected by the field nature of the specialism covered by the journal. Journal rankings within a field, such as JIF subject rankings in Clarivate's Journal Citation Reports, are more relevant but still subject to arbitrary variations by narrow specialism, calculation method and time. Low subject rankings or JIFs will never be used by managers as evidence that an article is low quality.

Interpreting indicators

Managers, appraisers and REF coordinators must consider time, field and career differences when evaluating any indicators presented by academics in support of their claims.

- The usefulness of citation indicators varies been fields and they are largely irrelevant in the arts and humanities. As a rough guide, managers should consult Table A3 of *Supplementary Report II: Correlation analysis of REF2014 scores and metrics* at http://www.hefce.ac.uk/pubs/rereports/year/2015/metrictide/.
- Average citation rates vary dramatically between fields. Citation counts, JIFs, h-indexes, career total citations should never be compared between different fields.
- Average citation rates vary between document types (e.g., journal articles, reviews, books, chapters) and should therefore not be compared between different document types.
- Average citation rates increase non-linearly over time and so managers should recognise that older articles are likely to be more cited than younger articles. Average citations per year is not a good substitute because of the non-linear accumulation pattern.
- Career-based indicators, such as total publication counts, total citation counts and the h-index are biased against females, due to their greater likelihood of career breaks for childcare or other carer responsibilities. They are also biased against people with temporary or permanent disabilities or illnesses, including all factor that counted as "special circumstances" in REF2014 that curtail their research productivities. Managers will make allowances for these factors when interpreting their value.
- The h-index should not be used because it conflates different types of research contribution.

VERSION:	3	AUTHOR/ OWNER:	Professor Mike Thelwall /Research Policy Unit
Approved Date:	May 2022	Approved By:	Academic Board
Review Date:	May 2025		

Related documents are copied below, for information.

DORA

General Recommendation
1) Do not use journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions.

For funding agencies
2) Be explicit about the criteria used in evaluating the scientific productivity of grant applicants and clearly highlight, 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.
3) 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.

For institutions
4) 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.
5) 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.

For publishers
6) Greatly reduce emphasis on the journal impact factor as a promotional tool, ideally by ceasing to promote the impact factor or by presenting the metric in the context of a variety of journal-based metrics (e.g., 5-year impact factor, EigenFactor [8], SCImago [9], h-index, editorial and publication times, etc.) that provide a richer view of journal performance.
7) Make available a range of article-level metrics to encourage a shift toward assessment based on the scientific content of an article rather than publication metrics of the journal in which it was published.
8) Encourage responsible authorship practices and the provision of information about the specific contributions of each author.
9) Whether a journal is open-access or subscription-based, remove all reuse limitations on reference lists in research articles and make them available under the Creative Commons Public Domain Dedication [10].
10) Remove or reduce the constraints on the number of references in research articles, and, where appropriate, mandate the citation of primary literature in favor of reviews in order to give credit to the group(s) who first reported a finding.

For organizations that supply metrics
11) Be open and transparent by providing data and methods used to calculate all metrics.
12) Provide the data under a licence that allows unrestricted reuse, and provide computational access to data, where possible.
13) Be clear that inappropriate manipulation of metrics will not be tolerated; be explicit about what constitutes inappropriate manipulation and what measures will be taken to combat this.
14) Account for the variation in article types (e.g., reviews versus research articles), and in different subject areas when metrics are used, aggregated, or compared.

For researchers
15) When involved in committees making decisions about funding, hiring, tenure, or promotion, make assessments based on scientific content rather than publication metrics.
16) Wherever appropriate, cite primary literature in which observations are first reported rather than reviews in order to give credit where credit is due.
17) Use a range of article metrics and indicators on personal/supporting statements, as evidence of the impact of individual published articles and other research outputs [11].

18) Challenge research assessment practices that rely inappropriately on Journal Impact Factors and promote and teach best practice that focuses on the value and influence of specific research outputs.

Metric tide (excerpt)

Responsible metrics can be understood in terms of the following dimensions:

- **Robustness**: basing metrics on the best possible data in terms of accuracy and scope
- **Humility**: recognising that quantitative evaluation should support – but not supplant – qualitative, expert assessment
- **Transparency**: keeping data collection and analytical processes open and transparent, so that those being evaluated can test and verify the results
- **Diversity**: accounting for variation by field, and using a range of indicators to reflect and support a plurality of research and researcher career paths across the system
- **Reflexivity**: recognising and anticipating the systemic and potential effects of indicators, and updating them in response.

Leiden Manifesto

Ten principles
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 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 employment[5]. 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 indicator[6]. 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.

"Simplicity is a virtue in an indicator because it enhances transparency."

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 averages[7].

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 (ref. 8). 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 Scholar[9]. 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].

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.