





SAGE was founded in 1965 by Sara Miller McCune to support the dissemination of usable knowledge by publishing innovative and high-quality research and teaching content. Today, we publish over 900 journals, including those of more than 400 learned societies, more than 800 new books per year, and a growing range of library products including archives, data, case studies, reports, and video. SAGE remains majority-owned by our founder, and after Sara's lifetime will become owned by a charitable trust that secures our continued independence.

Los Angeles | London | New Delhi | Singapore | Washington DC







THE INDEPENDENT REVIEW OF THE ROLE OF METRICS IN RESEARCH ASSESSMENT & MANAGEMENT

JAMES WILSDON

REPORT AUTHORS

JAMES WILSDON, LIZ ALLEN, ELEONORA BELFIORE, PHILIP CAMPBELL, STEPHEN CURRY, STEVEN HILL, RICHARD JONES, ROGER KAIN, SIMON KERRIDGE, MIKE THELWALL, JANE TINKLER, IAN VINEY, PAUL WOUTERS, JUDE HILL AND BEN JOHNSON







Los Angeles | London | New Delhi Singapore | Washington DC

SAGE Publications Ltd 1 Oliver's Yard 55 City Road London EC1Y 1SP

SAGE Publications Inc. 2455 Teller Road Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd B 1/I 1 Mohan Cooperative Industrial Area Mathura Road New Delhi 110 044

SAGE Publications Asia-Pacific Pte Ltd 3 Church Street #10-04 Samsung Hub Singapore 049483

Typeset by: C&M Digitals (P) Ltd, Chennai, India Printed and bound by CPI Group (UK) Ltd,

Introduction © James Wilsdon Report © HEFCE 2015, where indicated. The parts of this work that are @ HEFCE are available under the Open Government Licence 2.0: www.nationalarchives. gov.uk/doc/open-government-licence/version/2

This report was originally published in 2015 This version of the report published in 2016

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act, 1988, this publication may be reproduced, stored or transmitted in any form, or by any means, only with the prior permission in writing of the publishers, or in the case of reprographic reproduction, in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

Croydon, CR0 4YY

Library of Congress Control Number: 2015960254

British Library Cataloguing in Publication data

A catalogue record for this book is available from the British Library



ISBN 978-1-47397-306-0 (pbk)

At SAGE we take sustainability seriously. Most of our products are printed in the UK using FSC papers and boards. When we print overseas we ensure sustainable papers are used as measured by the PREPS grading system. We undertake an annual audit to monitor our sustainability.







ONE MEASURING UP

CHAPTER CONTENTS

1.1	Our terms of reference	. 3
1.2	Definitions and terminology	. 4
1.3	Data collection and analysis	. 7
1.4	The structure of this report	10

"The standing of British science, and the individuals and institutions that comprise it, is rooted firmly in excellence... Much of the confidence in standards of excellence promoted comes from decisions being informed by peer-review: leading experts assessing the quality of proposals and work."

Our Plan for Growth: science and innovation, HM Treasury/BIS, December 2014¹

"We have more top ranking universities in London than in any other city in the world. With 4 universities in the global top 10, we rank second only to the US."

Jo Johnson MP, Minister for Universities and Science, 1 June 2015²

¹https://www.gov.uk/government/publications/our-plan-for-growth-science-and-innovation

²Speech to 'Going Global' 2015 conference https://www.gov.uk/government/speeches/international-higher-education







Citations, journal impact factors, h-indices, even tweets and Facebook likes – there are no end of quantitative measures that can now be used to try to assess the quality and wider impacts of research. But how robust and reliable are such metrics, and what weight – if any – should we give them in the future management of research systems at the national or institutional level?

These are questions that have been explored over the past year by the Independent Review of the Role of Metrics in Research Assessment. The review was announced by David Willetts, then Minister for Universities and Science, in April 2014, and has been supported by the Higher Education Funding Council for England (HEFCE).

As the 2014 BIS/HM Treasury science and innovation strategy reminds us, the UK has a remarkable breadth of excellent research across the sciences, engineering, social sciences, arts and humanities. These strengths are often expressed in metric shorthand: "with just 3% of global research spending, 0.9% of global population and 4.1% of the world's researchers, the UK produces 9.5% of article downloads, 11.6% of citations and 15.9% of the world's most highly-cited articles".³

The quality and productivity of our research base is, at least in part, the result of smart management of the dual-support system of research funding. Since the introduction of the Research Assessment Exercise (RAE) in 1986, the UK has been through six cycles of evaluation and assessment, the latest of which was the 2014 Research Excellence Framework (REF2014). Processes to ensure and improve research quality, and more recently its wider impacts, are also used by the UK Research Councils, by other funders such as the Wellcome Trust, and by universities themselves.

The quality and diverse impacts of research have traditionally been assessed using a combination of peer review and a variety of quantitative indicators. Peer review has long been the most widely used method, and underpins the academic system in the UK and around the world. The use of metrics is a newer approach, but has developed rapidly over the past 20 years as a potential method of measuring research quality and impact in some fields. How best to do this remains the subject of considerable debate.

There are powerful currents whipping up the metric tide. These include growing pressures for audit and evaluation of public spending on higher education and research; demands by policymakers for more strategic intelligence on research quality and impact; the need for institutions to manage and develop their strategies for research; competition within and between institutions for prestige, students, staff and resources; and increases in the availability of real-time 'big data' on research uptake, and the capacity of tools for analysing them.

In a positive sense, wider use of quantitative indicators, and the emergence of alternative metrics for societal impact, can be seen as part of the transition to a more open, accountable





³Elsevier. (2013). International Comparative Performance of the UK Research Base – 2013; A report prepared by Elsevier for the UK's Department of Business, Innovation and Skills (BIS), p2. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/263729/bis-13-1297-international-comparative-performance-of-the-UK-research-base-2013.pdf. Retrieved 1 May 2015.



and outward-facing research system. But this has been accompanied by a backlash against the inappropriate weight being placed on particular indicators – such as journal impact factors (JIFs) – within the research system, as reflected by the 2013 San Francisco Declaration on Research Assessment (DORA), which now has over 570 organisational and 12,300 individual signatories. As DORA argues, "The outputs from scientific research are many and varied... Funding agencies, institutions that employ scientists, and scientists themselves, all have a desire, and need, to assess the quality and impact of scientific outputs. It is thus imperative that scientific output is measured accurately and evaluated wisely."

1.1 OUR TERMS OF REFERENCE

Our work builds on an earlier pilot exercise in 2008 and 2009, which tested the potential for using bibliometric indicators of research quality in REF2014. At that time, it was concluded that citation information was insufficiently robust to be used formulaically or as a primary indicator of quality, but that there might be scope for it to enhance processes of expert review.

This review has gone beyond the earlier pilot study to take a deeper and broader look at the potential uses and limitations of research metrics and indicators. It has explored the use of metrics across different disciplines, and assessed their potential contribution to the development of research excellence and impact within higher education. It has also analysed their role in processes of research assessment, including the next cycle of the REF. And it has considered the changing ways in which universities are using metrics, particularly the growing power of league tables and rankings. Finally, it has considered the relationship between the use of indicators and issues of equality and diversity, and the potential for 'gaming' that can arise from the use of particular indicators in systems of funding and evaluation.

To give structure and focus to our efforts, clear terms of reference were established at the outset. The review was asked to examine:

- The relative merits of different metrics in assessing the academic qualities and diverse impacts of research;
- The advantages and disadvantages of using metrics, compared with peer review, in creating an environment that enables and encourages excellent research and diverse impact, including fostering inter- and multidisciplinary research;





⁴Royal Society. (2012). *Science as an Open Enterprise*. The Royal Society Science Policy Centre report 02/12 https://royalsociety.org/~/media/policy/projects/sape/2012-06-20-saoe.pdf. Retrieved 1 June 2015.

^{*}www.ascb.org/dora. As of June 2015, only three UK universities are DORA signatories: Manchester, Sussex and UCL.

⁶Ibid.



- How metrics-based research assessment fits within the missions of universities and
 research institutes, and the value that they place on published research outputs in
 relation to the portfolio of other activities undertaken by their staff, including training
 and education;
- The appropriate balance between peer review and metrics in research assessment, and the consequences of shifting that balance for administrative burden and research cultures across different disciplines;
- What is not, or cannot, be measured by quantitative metrics;
- The differential impacts of metrics-based assessment on individual researchers, including the implications for early-career researchers, equality and diversity;
- Ethical considerations, and guidance on how to reduce the unintended effects and inappropriate use of metrics and university league-tables, including the impact of metrics-based assessment on research culture;
- The extent to which metrics could be used in novel ways by higher education institutions (HEIs) and research funders to support the assessment and management of research;
- The potential contribution of metrics to other aspects of research assessment, such as the matching of reviewers to proposals, or research portfolio analysis;
- The use of metrics in broader aspects of government science, innovation and industrial policy.

Reflecting the evidence we received, this report focuses in greater depth on some aspects of these terms of reference than others (notably, the use of metrics in the REF, by other funders and in HEI management). However, we hope that the report provides a clear framework for thinking about the broader role of metrics, data and indicators within research management, and lays helpful foundations for further work to be carried out by HEFCE, the Research Councils and others.

The review has been conducted in an open and consultative manner, with the aim of drawing in evidence, views and perspectives from across the higher education and research system. There has been a strong emphasis on transparency and plurality throughout the project, and the make-up of the review's steering group itself reflects a diversity of disciplines and perspectives. In addition, the group has engaged actively with stakeholders from across the research community through numerous workshops, meetings, talks and other channels, including the review's website and social media. Papers from steering group meetings have been made publicly available at every stage, as have other resources, including evidence received and slides presented at workshops.⁷

1.2 DEFINITIONS AND TERMINOLOGY

The research assessment landscape is contested, contentious and complex. Researchers, funders and managers face an ever-expanding menu of indicators, metrics and assessment methods in operation, many of which are explored in this review. Some are founded on





All of this material is available at the review's website: https://www.hefce.ac.uk/rsrch/metrics/



peer review, others on quantitative indicators such as citation counts, or measures of input, such as research funding or student numbers.

The term 'metric' is itself open to misunderstanding, because something can be a metric in one context but not in another. For example, the number of citations received by a researcher's publications is a citation metric but not an impact metric because it does not directly measure the impact of that researcher's work. In other words, it can imply 'measurement' of a quantity or quality which has not in fact been measured. The term *indicator* is preferable in contexts in which there is the potential for confusion. To reduce the scope of possible misunderstanding, this report will adopt the following definitions and terminology throughout.

Inc	dicators	A measurable quantity that 'stands in' or substitutes for something less readily measurable and is presumed to associate with it without directly measuring it. For example, citation counts could be used as indicators for the scientific impact of journal articles even though scientific impacts can occur in ways that do not generate citations. Similarly, counts of online syllabi mentioning a particular book might be used as an indicator of its educational impact.
Bil	bliometrics	Bibliometrics focuses on the quantitative analysis of scientific and scholarly publications, including patents. Bibliometrics is part of the field of scientometrics: the measurement of all aspects of science and technology, which may encompass information about any kind of research output (data, reagents, software, researcher interactions, funding, research commercialisation, and other outputs).8
Cit	tation	The most widely exploited bibliometric relies on counts of citations. Citation counts are
im	pact	sometimes used as an indicator of academic impact in the sense that citations from other documents suggest that the cited work has influenced the citing work in some way. Bibliometric indicators might normalise these citation counts by research field and by year, to take into account the very different citation behaviours between disciplines and the increase in citations over time. It has to be emphasised that as bibliometrics often do not distinguish between negative or positive citation, highly cited literature might attract attention due to controversy or even error. High numbers of citations might also result from a range of different contributions to a field e.g. including papers that establish new
		methodologies or systematically review the field, as well as primary research articles.
	ternative	Altmetrics are non-traditional metrics that cover not just citation counts but also
or	altmetrics	downloads, social media shares and other measures of impact of research outputs. The term is variously used to mean 'alternative metrics' or 'article level metrics', and it encompasses webometrics, or cybermetrics, which measure the features and relationships of online items, such as websites and log files. The rise of new social media has created an additional stream of work under the label altmetrics. These are indicators derived from social websites, such as Twitter, Academia.edu, Mendeley, and
		ResearchGate with data that can be gathered automatically by computer programs.

(Continued)





⁸Definitions adapted from *Encyclopedia of Science Technology and Ethics,* 2nd Edition (2014). Macmillan.



(Continued)

Peer review	A process of research assessment based on the use of expert deliberation and judgement. 9
Academic	Academic or scholarly impact is a recorded or otherwise auditable occasion of
or scholarly	influence from academic research on another researcher, university organisation or
impact	academic author. Academic impacts are most objectively demonstrated by citation
	indicators in those fields that publish in international journals. ¹⁰
Societal	As for academic or scholarly impact, though where the effect or influence reaches
impact	beyond scholarly research, e.g. on education, society, culture or the economy.
	Research has a societal impact when auditable or recorded influence is achieved upon
	$non-acade mic\ organisation (s)\ or\ actor (s)\ in\ a\ sector\ outside\ the\ university\ sector\ itself-for$
	instance, by being used by one or more business corporations, government bodies, civil
	society organisations, media or specialist/professional media organisations or in public
	debate. As is the case with academic impacts, societal impacts need to be demonstrated
	ratherthanassumed.Evidenceofexternalimpactscantaketheformofreferencesto,
	citations of or discussion of a person, their work or research results.11
REF impact	For the purposes of the REF2014,12 impact was defined as an effect on, change
	or benefit to the economy, society, culture, public policy or services, health, the
	environment or quality of life, beyond academia. REF2014 impact includes, but was not
	limited to, an effect on, change or benefit to:
	the activity, attitude, awareness, behaviour, capacity, opportunity, performance,
	policy, practice, process or understanding
	of an audience, beneficiary, community, constituency, organisation or individuals
	in any geographic location whether locally, regionally, nationally or internationally.
REF	Within REF2014, the research environment was assessed in terms of its 'vitality and
environment	sustainability', including its contribution to the vitality and sustainability of the wider
	discipline or research base.
REF outputs	Within REF2014, panels assessed the quality of submitted research outputs in terms of
	their 'originality, significance and rigour', with reference to international research quality
	standards. ¹³

⁹Adapted from: Council of Canadian Academies. (2012). *Informing Research Choices: Indicators and Judgment*, p11. www.scienceadvice.ca/uploads/eng/assessments%20and%20publications%20and%20news%20releases/science%20performance/science performance_fullreport_en_web.pdf. Retrieved 6 December 2014.

¹⁰Taken from LSE Public Policy Group (2011) *Maximising the Impacts of Your Research: A Handbook for Social Scientists*. London: PPG. http://blogs.lse.ac.uk/impactofsocialsciences/the-handbook/.

11Ibid.

¹²REF 02. 2011. Assessment framework and guidance on submissions, p26, para 141. www.ref.ac.uk/media/ref/content/pub/assessmentframeworkandguidanceonsubmissions/GOS%20including%20addendum.pdf. Retrieved 2 April 2015.

¹³Ibid, p23, para 118, notes that permitted 'types' of outputs included: Books (or parts of books); Journal articles and conference contributions; Physical artefacts; Exhibitions and performances; Other documents; Digital artefacts (including web content); Other.







1.3 DATA COLLECTION AND ANALYSIS

The review drew on an extensive range of evidence sources, including:

1.3.1 A formal call for evidence

A call for evidence was launched on 1 May 2014, with a response deadline of 30 June 2014. The steering group appealed for evidence from a wide range of sources, including written summaries or published research. Respondents were asked to focus on four key themes and associated questions, as follows:

- A Identifying useful metrics for research assessment.
- B How metrics should be used in research assessment.
- C 'Gaming' and strategic use of metrics.
- D International perspective.

In total, 153 responses were received to the call for evidence: 67 from HEIs, 42 from individuals, 27 from learned societies, 11 from publishers and data providers, three from HE mission groups, and three from other respondents. An analysis of the evidence received can be found at www.hefce.ac.uk/rsrch/metrics/call/.

1.3.2 A literature review

Two members of the Steering Group, Paul Wouters and Michael Thelwall, researched and wrote a comprehensive literature review to inform the review's work. The findings of the literature review have been incorporated into this report at appropriate points, and the full review is available as Supplementary Report I.¹⁵

1.3.3 Community and stakeholder engagement

The review team engaged actively with stakeholders across the higher education and research community. These activities included a series of six workshops, organised by the steering group, on specific aspects of the review, such as the role of metrics within the arts and humanities, and links to equality and diversity. Members of the steering group also gave talks and presentations about the work of the review at around 30 conferences, roundtables and workshops. Findings and insights from these events have







¹⁴The call for evidence letter is available at: www.hefce.ac.uk/media/hefce/content/What,we,do/Research/How,we,fund,research/Metrics/Letter-call-for-evidence-metrics-review.pdf

¹⁵Wouters, P., et al. (2015). *Literature Review: Supplementary Report to the Independent Review of the Role of Metrics in Research Assessment and Management*. HEFCE. DOI: 10.13140/RG.2.1.5066.3520.



been incorporated into the report wherever appropriate. A full itinerary of events linked to the review can be found in the 'Annex of tables' at the end of this report (Table 2).

1.3.4 Media and social media

Over the course of the review, the steering group sought to encourage wider discussion of these issues in the sector press (particularly Times Higher Education and Research Fortnight) and through social media. There was extensive use of the #HEFCEmetrics hashtag on Twitter. Members of the steering group, including Stephen Curry, ¹⁶ also wrote blog posts on issues relating to the review, and a number of other blog posts and articles were written in response to the review. ¹⁷

1.3.5 Focus groups with REF2014 panel members

The steering group participated in a series of focus group sessions for REF2014 panel members, organised by HEFCE, to allow panellists to reflect on their experience, and wider strengths and weaknesses of the exercise. Specific sessions explored the pros and cons of any uses of metrics within REF2014, and their potential role in future assessment exercises.

1.3.6 REF2014 evaluations

Where relevant, the steering group also engaged with and analysed findings from HEFCE's portfolio of REF2014 evaluation projects, including:

 The nature, scale and beneficiaries of research impact: an initial analysis of REF2014 case studies;¹⁸

¹6Curry, S. (2014). Debating the role of metrics in research assessment. Blog posted at http://occamstypewriter.org/scurry/2014/10/07/debating-the-role-of-metrics-in-research-assessment/. Retrieved 1 June 2015.

¹⁷Numerous blog posts, including contributions from steering group members, have been featured at http://blogs.lse.ac.uk/impactofsocialsciences/2014/04/03/reading-list-for-hef cemetrics/. Retrieved 1 June 2015. We have referred to some of these posts within this report. Others discussing the review through blog posts include: David Colquhoun, www.dcscience.net/2014/06/18/should-metrics-be-used-to-assess-research-performance-a-submission-to-hefce/. Retrieved 1 June 2015. Also see contributors to: http://thedisorderofthings.com/tag/metrics/. Retrieved 1 June 2015.

¹⁸King's College London and Digital Science. (2015). *The Nature, Scale and Beneficiaries of Research Impact: An Initial Analysis of Research Excellence Framework (REF) 2014 impact case studies.* www.hefce.ac.uk/pubs/rereports/Year/2015/analysisREFimpact/. Retrieved 1 June 2015.







- Preparing impact submissions for REF2014;¹⁹
- Assessing impact submissions for REF2014;²⁰
- Evaluating the 2014 REF: Feedback from participating institutions;²¹
- REF Manager's report;²²
- REF panel overview reports;²³
- REF Accountability Review: costs, benefits and burden project report.²⁴

1.3.7 Relating REF2014 outcomes to indicators

A final element of our evidence gathering was designed to assess the extent to which the outcome of the REF2014 assessment correlated with 15 metrics-based indicators of research performance. For the first time, we were able to associate anonymised REF authors by paper outputs to a selection of metric indicators, including ten bibliometric indicators and five alternative metric indicators. Previous research in this area has been restricted to specific subject areas and departmental level metrics, as the detailed level of data required for this analysis was destroyed before publication of the REF2014 results. This work is summarised in Chapter 9, and presented in detail in Supplementary Report II.²⁵





¹⁹Manville, C., Morgan Jones, M, Frearson, M., Castle-Clarke, S., Henham, M., Gunashekar, S. and Grant, J. (2015). *Preparing Impact Submissions for REF2014: Findings and Observations*. Santa Monica, CA: RAND Corporation. RR-727-HEFCE. www.hefce.ac.uk/media/HEFCE,2014/Content/Pubs/Independentresearch/2015/REF,impact,submissions/REF_impact_prep_process-findings.pdf

²⁰Manville, C., Guthrie, S., Henham, M., Garrod, B., Sousa, S., Kirtley, A., Castle-Clark, S. and Ling, T. (2015). *Assessing Impact Submissions for REF2014: An Evaluation*. Santa Monica, CA: RAND Corp. www.hefce.ac.uk/media/HEFCE,2014/Content/Pubs/Independentresearch/2015/REF,impact,submissions/REF_assessing_impact_submissions.pdf

²¹HEFCE. (2015). Evaluating the 2014 REF: Feedback from Participating Institutions. www.hefce. ac.uk/media/HEFCE,2014/Content/Research/Review,of,REF/2014_REF_sector_feedback. pdf

²²HEFCE. (2015). Research Excellence Framework 2014: Manager's Report. www.ref.ac.uk/media/ref/content/pub/REF_managers_report.pdf. Retrieved 25 May 2015

²³HEFCE's Panel overview reports can be downloaded from www.ref.ac.uk/panels/panelover viewreports/

²⁴Technopolis, 2015.

²⁵HEFCE (2015). Correlation analysis of REF2014 scores and metrics: *Supplementary Report II to the Independent Review of the Role of Metrics in Research Assessment and Management*. HEFCE. DOI: 10.13140/RG.2.1.3362.4162.



1.4 THE STRUCTURE OF THIS REPORT

This opening chapter has provided a summary of the aims and working methods of the review, and the range of evidence sources on which this final report draws.

Chapter 2 (*The rising tide*) gives a brief history of the role of metrics in research management, and the evolution of data infrastructure and standards to underpin more complex and varied uses of quantitative indicators. It also surveys the main features of research assessment systems in a handful of countries: Australia, Denmark, Italy, the Netherlands, New Zealand and the United States.

Chapter 3 (*Rough indications*) looks in greater detail at the development, uses and occasional abuses of four categories of quantitative indicators: bibliometric indicators of research quality; alternative indicators of quality; input indicators; and indicators of impact.

Chapter 4 (*Disciplinary dilemmas*) maps the diversity in types of research output, publication practices and citation cultures across different disciplines, and the implications these have for any attempts to develop standardised indicators across the entire research base. It also considers the extent to which quantitative indicators can be used to support or suppress multi- or interdisciplinary research.

Chapter 5 (*Judgement and peer review*) compares the strengths and weaknesses of the peer review system with metric-based alternatives, and asks how we strike an appropriate balance between quantitative indicators and expert judgement.

Chapter 6 (*Management by metrics*) charts the rise of more formal systems of research management within HEIs, and the growing significance that is being placed on quantitative indicators, both within institutions and as a way of benchmarking performance against others. It looks specifically at university rankings and league tables as a visible manifestation of these trends, and considers how these might be applied in more responsible ways across the sector.

Chapter 7 (*Cultures of counting*) assesses the wider effects a heightened emphasis on quantitative indicators may have on cultures and practices of research, including concerns over systems for performance management, and negative effects on interdisciplinarity, equality and diversity. It also considers the extent to which metrics exacerbate problems of gaming and strategic approaches to research assessment.

Chapter 8 (*Sciences in transition*) looks beyond HEIs to examine changes in the way key institutions in the wider research funding system are using quantitative indicators, including the Research Councils, research charities such as the Wellcome Trust, and the national academies. It also looks to developments at the European level, within Horizon2020. Finally, it considers how government could make greater use of available quantitative data sources to inform horizon scanning and policies for research and innovation.

Chapter 9 (*Reflections on REF*) provides a detailed analysis of the modest role that quantitative indicators played in REF2014, and considers a range of scenarios for their use in future assessment exercises. It also outlines the results of our own quantitative







analysis, which correlated the actual outcomes of REF2014 against 15 metrics-based indicators of research performance.

Finally, Chapter 10 (*Responsible metrics*) summarises our headline findings, and makes a set of targeted recommendations to HEIs, research funders (including HEFCE), publishers and data providers, government and the wider research community. Within a framework of *responsible metrics*, the report concludes with clear guidance on how quantitative indicators can be used intelligently and appropriately to further strengthen the quality and impacts of UK research.



