

Independent review of the role of metrics in research assessment

University Alliance response, June 2014

1. University Alliance is a non-partisan, non-political organisation working to promote, safeguard and sustain the public benefit delivered by universities.
2. University Alliance brings together 22 of the UK's leading innovative and enterprising universities – major institutions combining science, technology and the creative industries with a focus on delivering for the professions, business and the community. Alliance universities have over 300 research teams undertaking world-leading research, and employ 1 in 6 of the UK's university researchers. Alliance universities support around a third of all UK research teams undertaking world class research in Allied Health and over a quarter of those in General Engineering, Architecture and the Built Environment, drawing over 50% of their research income in STEM-related areas. They lead the enterprise and entrepreneurship agenda with over 21,000 business links (including 14,000 SMEs) and generate nearly 50% of all turnover from UK graduate start-ups.

Summary

3. Quantitative metrics are used to measure vastly different aspects of research. Hence draws a distinction between 'quality' and 'impact' in consideration of research and for the parameters of this review. Using these as broad, if sometimes overlapping categories, it is possible to identify metrics which reflect the reach of research within the academic community (citations/bibliometrics-based measures), usually held to indicate 'quality', and those which seek to measure its impact in other ways, beyond scholarly debate (these might include, amongst many others, YouTube hits, social media shares, etc.). We consider these categories of metrics-based systems separately here as 'scholarly' and 'societal' measures,¹ since the terms 'quality' and 'impact' can confuse when both are, necessarily, intertwined – excellent research should be relevant and have impact on the scholarly field as well as beyond. This highlights the need to distinguish clearly the aim of research being assessed, striking the balance between 'scholarly' and 'societal' – that is to say, the balance between the academic reach of research and its impact more widely.
4. Insofar as metrics may have implications for research funding allocation, it is essential that the principle of funding excellence wherever it exists – for which the evidence is

¹ Here (and throughout) 'societal' encapsulates all impacts – including economic impacts – which exist outside of scholarly debate.

well established² - is upheld, as an important pillar of the UK's dual funding system. The UK must continue its policy of selectively distributing research funding, based on quality, in order to continue to drive the quality and impact of UK research and secure the future health of the UK research base. In a difficult fiscal environment it is essential that these existing principles are maintained because they have "enabled the Government and funding bodies to maximise the return from the limited public funds available for ...research".³

5. In the context of current fiscal pressures, it is right that we should consider ways to reduce wastefulness and inefficiencies in the system – for example around allocation and application methods for research funding – as long as the underpinning principals of seeking and supporting excellence in the system are maintained. We should be careful that this approach is not threatened as a result of limiting assessment to only that which is measurable by metrics. Peer review ensures that areas of strength (and weakness) are recognised as such, rather than overall metrics which cannot recognise nuance. The risk is an over-simplified view of the research ecosystem which – if amplified by funding decisions – risks weakening the whole system, through narrowing its scope and reducing capacity.
6. In an increasingly digitised age for research publications, and with the advent of open access requirements, it is timely to re-appraise the role of metrics in research assessment for written, scholarly outputs. It may be too early to appraise fully the role of metrics for the measurement of societal impact when this concept is still relatively new as a measurable aspect of research, although metrics may help to standardise impacts across units of assessment. However, there remain concerns that metrics-based systems can amplify existing prejudices, and reinforce inherent conservatism in the system, thereby restricting innovation in scholarship.
7. Until these concerns can be fully and robustly addressed, the use of metrics as a primary tool in assessing research activities should be avoided, particularly when there are funding implications which may further extend these prejudices in the future.
8. There is also a question of scale – since compelling evidence shows that quality is a driver of scale and not vice-versa⁴ – therefore metrics which may be influenced by scale (including citations and economic measurements) would not be indicative of quality. The goal of national research assessment mechanisms (and related performance management aims) is surely to foster an excellent research ecosystem. Efficiency and productivity are important, therefore if a recommendation is made to integrate metrics

² [L. Aston and L. Shutt \(2009\). *Concentration and diversity: understanding the relationship between excellence, concentration and critical mass in UK research* and *University Alliance \(2011\). Funding research excellence: research group size, critical mass & performance*](#)

³ www.rae.ac.uk/Pubs/2004/01/rae0401.doc

⁴ [Evidence Ltd \(2011\). *Funding research excellence: research group size, critical mass & performance*, University Alliance](#)

into UK research assessment mechanisms, it will be essential that these are scaled, taking account of research capacity.

9. There are also significant concerns about the ability to standardise metrics across disciplines. This manifests itself particularly in the divergences in publication culture between Arts & Humanities and STEM subjects, as outlined below. There are likely to be similar (if less evident) differences between the types of societal impact that different disciplines might have. An over-reliance on metrics-based assessment may result in too much deviation for a whole-system mechanism, and may discourage important cross-disciplinary work.

Identifying useful metrics for research assessment:

- **What empirical evidence (qualitative or quantitative) is needed for the evaluation of research, research outputs and career decisions?**
10. A serious analysis of the balance of outputs and impact is necessary to define the quality – as well as the unit – of research.
 11. REF 2014 assesses the ‘vitality and sustainability’ of the research environment. This must necessarily assess the environment in supporting researchers to realise societal benefit as well as ‘scholarly’ reach (this could include partnerships and collaborative research as well as general support for realising benefits of research more widely).
- **What metric indicators are currently useful for the assessment of research outputs, research impacts and research environments?**
12. **Research outputs/scholarly reach.** There is evidence that shows that “whilst citations are a very useful indicator at high levels of aggregation – for example at subject level, to assess the relative quality of UK research – they are increasingly unreliable at lower levels of analysis, such as the individual researcher, research group or department.”⁵
- **What new metrics, not readily available currently, might be useful in the future?**
13. **Scholarly metrics.** Alternative bibliometric approaches which make citation analysis more granular such as allows to make more nuanced analyses (i.e. agrees with, disagrees with, cites as authority, cites as data source) may be helpful. Metrics that are more accurate for publications in languages other than English would also be welcomed.
 14. **Societal metrics.** Research blogs, navigation of research through Idea Pod and other web resources, including institutional repositories, are potential mechanisms and resources for assessing quality through richer content capture methodologies, involving wider reach beyond academic circles of ideas as well as potential impacts.
 15. Metrics which seek to demonstrate collaboration with non-academic partners through co-authored papers may not be sufficient alone to demonstrate industry collaboration, particularly since much of the research that is useful to business does not result in such

⁵ [T. Sastry and Bahram Bekhradnia \(2006\). *Using metrics to allocate research funds*, HEPI.](#)

an output. These metrics must be complemented by other measures of university-business collaborative research including quantity of business links as measured by number of different income sources.

16. The REF2014 criteria for assessing impacts are ‘reach’ and ‘significance’. The process of assessing qualitative portrayals of whole research projects such as impact case studies submitted to REF2014 is currently underway. Metrics-based measurement of elements of impact may be beneficial - there may be a role for metrics-based analyses to help gauge ‘reach’, in particular. ‘Significance’ may also be measurable by metrics, but since the categories of impact are vastly different (from economic impacts on companies, to education) efforts to standardise across will require difficult value judgments.
17. There is no single recognised measurement tool for impact of knowledge exchange, and none fully capture the range of activities and levels at which KE operates or have the ability to be generalised across institutions and research systems.
18. **Economic metrics.** Metrics measuring research impact in economic terms are widely used to demonstrate societal impact, but these need to be nuanced, scaled and any multipliers involved need to be transparent.
19. An example can be seen with demonstrating impact of research activities with SMEs in current metrics systems. High levels of research and innovation activities with SMEs, which can be costly and resource-intensive, do not necessarily translate into high levels of income (with implications for HE-BCI results and, subsequently, HEIF). Some measure of levels of SME engagement in quantity – given that this is a policy priority at all levels – would be extremely useful and a strong indicator of how universities are actively engaging with LEP and local growth agendas.
20. A Social Return on Investment (SRoI) approach to include research strategy - the unique contribution of the research centre or institution – would be helpful to demonstrate the value added by institutional support.
21. The NCPPE *Through a Glass, Darkly* report suggests a new metric taking previous systems into account - ‘Socially modified economic valuation’ (SMEV) - but this is yet to be fully tested or adopted;⁶ likewise a proposal for a new coded CERIF-based scheme is not yet in use.⁷
22. **Non-economic metrics.** In demonstrating societal impact John Goddard suggests using the Finnish, and subsequently OECD, approach of a standard report ‘covering the impact of the universities on economic, social, cultural and environmental development’ which is subsequently peer-reviewed.⁸
23. Current methods of measuring change of opinion and user perceptions (usually testimonies and surveys) are time-consuming and difficult to apply retrospectively:

⁶ NCCPE, *Through a Glass, Darkly: Measuring the Social Value of Universities*, sections 7-9.

⁷ Richard Gartner, Mark Cox, and Keith Jeffery, “A CERIF-Based Schema for Recording Research Impact,” *The Electronic Library* 31, no. 4 (2013): 465–482, doi:10.1108/EL-11-2011-0156.

⁸ John Goddard (2009). *Reinventing the Civic University*. NESTA.

developments in data-mining tools may allow interactions taking place between researchers, institutions, and stakeholders, to be electronically or automatically captured in the future. Capturing knowledge exchange events would help link research with impact - Penfield et al. highlight the Social Impact Assessment Methods through the study of Productive Interactions (SIAMPI), but SIAMPI was developed as a learning tool to understand more clearly how research interactions lead to social impact, rather than as a framework of judging impact and it cannot be used to comprehensively assess impact in a standardised manner.⁹

24. Metrics for knowledge exchange should be developed in collaboration with end-users (to develop how research leads to the development of products, concepts or interventions will better reflect the value of research impact for society). Evidence suggests that in one field of possible impact – knowledge exchange with business – current metrics are not useful. Businesses use a variety of methods to assess the impact of knowledge exchange activities. CIHE research showed that, beyond specific deliverables, businesses valued KE activities because of:
- a. the expertise developed in individual researchers as a result of a collection of previous research projects;
 - b. access to a broad base of information which includes research journals and conference papers to locate academics and understand their expertise;
 - c. the methods, techniques and data developed in the course of any research.
25. It is often therefore not possible to identify a direct ‘line of sight’ between research and a commercial result because of the long-running and complex relationships between new and existing ideas (both external and in-house), and business development processes, thus even in commercial contexts economic and metrics-based measures of impact are not sufficient. In fact, their research suggested that ‘other than in a very limited number of cases, (e.g. at the delivery stages) numerical metrics are likely to be inappropriate or misleading as an indicator of impact... Hence impact indicators should be appropriate to the stage of research it relates to, e.g. ranging from identifying future academic or industrial partners for application development through to the successful completion of a prototype with an industrial collaborator. Impact is identified as a response in potential user behaviour enabled and encouraged by research outcomes, not as a standard metric (such as license income or patents licensed).’¹⁰
- **Are there aspects of metrics that could be applied to research from different disciplines?**
26. Cost/benefit analysis and a formal Social Return on Investment (SRoI) approach might offer a meaningful starting point for research quality measurement in the round and may be more rigorous and comprehensive than presently, where output and outcome only are the foci of measurement. Some elements of societal impact assessment will be

⁹ T. Penfield et al. (October 2013), “Assessment, Evaluations, and Definitions of Research Impact: A Review,” *Research Evaluation* doi:10.1093/reseval/rvt021.

¹⁰ CIHE, *Absorbing Research: The Role of University Research in Business and Market Innovation*.

transferable across disciplines. Online impact portfolios would reflect more open and comprehensive rich data provision for service evaluation, as in other sectors.

- **What are the implications of the disciplinary differences in practices and norms of research culture for the use of metrics?**
27. Metrics are more helpful in analysing the academic reach of research in subjects where articles are the main form of output, but not as useful in other disciplines with different publication cultures, such as the Humanities, where monographs (equivalent to several long articles) are more common and where reach within the discipline spreads over longer timescales. It may take two or three years for a monograph to be reviewed, whereas scholarly reach of articles published in some STEM subjects may be measurable within two years. Arts and humanities subjects can also operate exclusively outside of journal impact factor-type measurements of quality, owing to their non- journal based outputs. ([Plumanalytics](#) cites example metrics for non-standard outputs.)
- **What are the best sources for bibliometric data? What evidence supports the reliability of these sources?**
28. No single bibliometric measure is perfect and there is evidence for large discrepancies – up to 160% - between large bibliometric providers (Scopus, Web of Science and Google Scholar, etc).¹¹ As research naturally occupies more of the space available to it in Web 2.0, these measures may not necessarily be assumed to be the sole or best future sources of research output quality.
- **What evidence supports the use of metrics as good indicators of research quality?**
29. The various methods of measuring dissemination and uptake within the research community (h-index, c-index, JIF, etc) are generally considered to be the most effective metrics-based mechanism for evaluation of ‘scholarly’ reach. Peer reviewed journals retain their lead as the preferred vehicle for formal measurement of research quality. There are serious concerns, however (outlined below), that these mechanisms are less than robust, are likely to exacerbate systemic prejudices and are vulnerable to gaming.
- **Is there evidence for the move to more open access to the research literature to enable new metrics to be used or enhance the usefulness of existing metrics?**
30. Heavy citation/bibliometric weighting would encourage researchers to focus publishing efforts in mainstream or market leading journals above and beyond other considerations. Open access, open writing, open data and open peer review processes offer significant opportunities to encourage diversity in quality research and to ensure that researchers are not encouraged to remain conservative in their research, especially early in their careers. A new range of journals with open ethos may also transform the use and engagement with research, with increased accessibility increasing the likelihood of societal impacts. Current dominance and hierarchies within journals may become less important, which in an expanded and democratic system may be no bad thing.

¹¹ Lokman I Meho (January, 2007), ‘The rise and rise of citation analysis’, *Physics World*, pp. 32-36.

How should metrics be used in research assessment?

31. Multiple metric and review approaches to evaluation of quality are recommended - in line with the San Francisco Declaration on Research Assessment (DORA), journal-based metrics should not be used in isolation due to inherent discrepancies and their overly-limited view in relation to the wider impact of research.
- **What examples are there of the use of metrics in research assessment?**
32. Peer review of outputs currently forms the main part of the REF although some disciplinary units utilise JIF to inform their quality evaluation. Metrics are used in relation to funding and PGR. Impact is a narrative assessment with limited required metrics. Some panels have reported the use of metrics in some cases.
33. Benchmarking software is available for institutions to measure their citation and JIF peer to peer.
- **To what extent is it possible to use metrics to capture the quality and significance of research?**
34. REF2014 guidance asks its panel members to apply criteria of 'originality, significance and rigour' to assessing research outputs. Citation analysis is not able to shed light on originality or rigour, but arguably it can measure 'significance' in terms of reach within the academic community only. Qualitative evidence – i.e. the research outputs themselves – relies on a (ideally 'blind') peer review system which may be better placed to recognise to measure 'originality' and 'rigour'. Quantitative metrics must therefore remain a complementary but not primary element of research assessment, to enhance but not replace the peer review system.
35. Alternative metrics as described in this response could be added to obtain a richer view of research quality, although it is important that any evaluation framework for research, whether metrics-based or output review by panels, was future-proofed to take account of new technologies and methodologies for research management.
- **Are there disciplines in which metrics could usefully play a greater or lesser role? What evidence is there to support or refute this?**
36. Traditional citation measures are less well-suited to assessing the quality of non-traditional content forms and certain disciplines. We would urge, however, a whole-system response that allows cross-disciplinary analysis and collaboration, and therefore the use of metrics only in minor and supporting roles to peer review.
- **How does the level at which metrics are calculated (nation, institution, research unit, journal, individual) impact on their usefulness and robustness?**

37. Citations a useful indicator at high levels of aggregation (i.e. at subject or nation level) but are increasingly unreliable at lower levels of analysis, such as the individual researcher or research unit.¹²
38. Analyses of research quality must be at disciplinary, rather than institutional, level in order to be meaningful. It should also be scaled according to research staff, not student body to represent return on research investment.
39. Individual researchers and institutions may change behaviours in relation to metrics-based assessment processes. Some concerns relating to gaming (and therefore robustness) are outlined below.

‘Gaming’ and strategic use of metrics

- **What evidence exists around the strategic behaviour of researchers, research managers and publishers responding to specific metrics?**
40. It has been shown how easy it is to manipulate certain metrics systems.¹³ Gaming is an inevitable consequence if widespread use of journal-based metrics is adopted, since it creates both personal and institutional incentives for a researcher to get cited at all costs and for a systemic response aligning outputs to rigid targets. Citation clubs and self-citation will be unavoidable and rankings will be very difficult to keep ‘clean’. This reduction of citation for the sake of citation also runs the risk of leading to sensationalism rather than rigour and discipline. This will further undermine the objective of ‘quality’ and academic freedom may be compromised, resulting in less diversity or innovation in research.
 41. Insofar as metrics may be used for rankings, there is a raft of evidence describing behavioural impacts – not all positive. These are summarized in a recent report by the European University Association.¹⁴
 - **What are the risks that some groups within the academic community might be disproportionately disadvantaged by the use of metrics for research assessment and management?**
 42. Citation metrics clearly favour older or more established researchers. Recent outputs will have fewer citations, and citations owed to reputation, rather than the quality of a particular piece of research, are magnified. They are subject to self-promotion and rankings manipulation, a process in which established authors have an advantage. In addition, citation metrics do not take into account author rank within multi-authored articles, and therefore do not reflect the contributions of authors.

¹² T. Sastry and Bahram Bekhradnia (2006). *Using metrics to allocate research funds*, HEPI.

¹³ Phil Davis, (2012), ‘Gaming Google Scholars Citations: Made Simple and Easy’, *Scholarly Kitchen* blog. Available at: <http://scholarlykitchen.sspnet.org/2012/12/12/gaming-google-scholar-citations-made-simple-and-easy/>

¹⁴ European University Association (2013). *Global University Rankings and their Impact. Report II*, pp. 21-22.

43. Citation metrics may also compound biases of scale, since they are subject to advantages of reputation. Evidence shows that research quality is a driver of scale and not vice-versa,¹⁵ and metrics which may be influenced by scale would not be indicative of quality.
44. Evidence also shows an inherent conservatism in the systematic under-citation of women and minority groups,¹⁶ and similar biases in recruitment and promotion processes.¹⁷
45. A move to metrics runs the risk of embedding such biases into research assessment. Rather than a commitment to seeking truly innovative research, dependence on citation analyses alone would result in a reinforcement of current hegemonies (of age, gender and ethnicity). If these metrics are used to allocate research funding, the consequence will be a greater concentration of research funding, opportunity and prestige and a decreased diversity of the research base – both in terms of human capital and the nature and spread of the disciplines supported – which will weaken the future research ecosystem. Diversity matters in future-proofing the system.
- **What can be done to minimise ‘gaming’ and ensure the use of metrics is as objective and fit-for-purpose as possible?**
46. Metrics need to be as comprehensive and wide-ranging as possible and assess elements including the process, rigour, reach and impact of research. They must be corrected for any inherent prejudices.

International perspective:

47. Thomson Reuters are writing a White Paper on potential new impact metrics, which will be opened for further global consultation, the results of which will be an interesting and useful addition to this consultation, particularly around quality measurements relating to societal impact.¹⁸
48. Details of underlying metrics in a variety of global rankings systems, including many research indicators, are discussed in the EUA report on Global Rankings, which also has a useful bibliography.¹⁹

¹⁵ Evidence Ltd (2011). *Funding research excellence: research group size, critical mass & performance*, University Alliance

¹⁶ Daniel Maliniak, Ryan Powers and Barbara F. Walter (2013). The Gender Citation Gap in International Relations. *International Organization*, 67, pp 889-922. doi:10.1017/S0020818313000209. See open version here:<http://politicalviolenceataglance.files.wordpress.com/2013/03/the-gender-citation-gap-in-ir.pdf>

¹⁷ Ilana Yurkiewicz, (2012) ‘Study Shows Gender Bias in Science is Real. Here’s Why It Matters’, *Scientific American*, available at <http://blogs.scientificamerican.com/unofficial-prognosis/2012/09/23/study-shows-gender-bias-in-science-is-real-heres-why-it-matters/>; April Corrice (2009) ‘Unconscious Bias in Faculty and Leadership Recruitment: A Literature Review’, Association of American Medical Colleges, available at [http://www.hopkinsmedicine.org/diversity_cultural_competence/pdf/Unconscious%20Bias%20in%](http://www.hopkinsmedicine.org/diversity_cultural_competence/pdf/Unconscious%20Bias%20in%20)

¹⁸ <http://www.timeshighereducation.co.uk/world-university-rankings/news/global-consultation-begins-on-new-university-rankings-impact-metrics>

¹⁹ European University Association (2013). *Global University Rankings and their Impact. Report II*.