

International Advisory Group

Future Research Assessment Programme (FRAP)

Published June 2023

The Future Research Programme Assessment Board
Research England

3 June 2023

Report of the International Advisory Group

Dear FRAP Board,

Please find attached the report of the International Advisory Group (IAG) which I had the privilege of chairing over the past two years. The group met entirely virtually. The report details how we operated in two phases and with reference to the terms of reference we operated under.

As an international advisory group, we recognized that it would be inappropriate to get into the operational details of the REF and thus focused only on the principles that should underpin research assessment in a dual funding system. In doing so we make thirteen recommendations for your consideration.

Because research assessment in the UK is linked to both funding and institutional and academic reputations, it is inevitable that there will be diverse views of any evolution of the research assessment process.

The goal of research assessment must be to ensure a sustainable UK research system. It should continue to evolve but in doing so provide a source of research funding linked to research excellence; give accountability for this research funding; and provide system-wide and institutional data for performance enhancement at those levels. It is desirable that the system becomes more forward-looking and therefore assessment should not be based entirely on past performance. While research assessment is thus about much more than funding, the dual funding approach with block-grant research funding linked to research assessment, appropriately defined, is in the UK broader strategic interests.

Indeed as an incentive system, the REF should be seeking to ensure the future of UK research in a rapidly evolving environment. The focus on outputs has had some negative impacts on research culture and the research and innovation landscape as a whole. There are, of course, wider issues within the research sector (e.g. EDI and representation, a somewhat toxic research culture). The REF cannot be seen as the sole tool for fixing these issues but it has an important role to play alongside that of grant funders, academies, and the institutions and scholars themselves. The introduction of impact assessment in 2014 was step forward in culture change but a focus on institutional level performance would allow new ways to be developed to recognise and reward positive research culture.

While the REF is much more than a funding tool, it would be naïve not to recognize the linkage to funding. Block-grant research funding provides a critical component of the funding of research in the UK's higher education system albeit it that there are somewhat different approaches across the devolved administrations. Importantly this is the major way in which non-competitive (research grant) funding for research is available to universities. It is thus a vital strategic component of the UK's capacity to be at the forefront of the global research enterprise. However, it is important that it is used strategically and innovatively by the institutions that receive block-grant funding: our recommendations are aimed to shift the incentives in that direction.

Globally, it is now clear that a strong focus on traditional output measures lead to some behaviours at institutional, disciplinary, and individual level that impede innovation and evolution of the research profile. Rather the IAG saw the need for a far greater obligation and incentives being placed on the institutional level of assessment, recognizing further work is needed on how best to assess that. We may have even been conservative in our recommendation regarding the reduced weighting of outputs and the enhancement of the institutional level assessment. But we also recognized the pragmatic reality that evolution rather than revolution was more likely to be acceptable.

We note that Sir Paul Nurse's recent review of UK research identified the need to make the use of block-grant funding more transparent. There is a logic to this that in turn links to a greater weighting towards institutional level of assessment. However, because this was beyond our terms of reference and there are undoubtedly complexities we could not evaluate, we did not make a specific recommendation in this regard.

The relationship between research assessment and funding has led to unrealistic expectations as to the precision of research assessment. Terms like quality, excellence and impact will always have some level of imprecision and this should be accepted rather than pretending that excess process can create a precision that cannot exist. It has been this demand for precision and the incentives thus created by a high weighting of outputs, that have driven the excessive focus on traditional outputs at the expense of recognizing the evolution of research in ways not well picked up by such methods. We are concerned that cross disciplinary and particularly transdisciplinary research (which by its definition if the processes are well conducted is impactful) needs incentivization. By reducing the weighting on outputs, it is hoped the burden of research assessment could over time be simplified.

The cost of the REF is clearly a concern. Beyond the matter of a likely over-engineered output assessment, the other major cost driver are the institutions themselves using the research assessment exercise not always appropriately in their own interests. We make comments on this challenge and encourage both the funding bodies and institutions to work together to reduce the cost drivers.

While we make some comments on AI and in particular the emergence of AI driven research processes and now of AI and large language model production of research outputs, this field is moving very rapidly. Only now is the research policy community turning its mind to the

underlying ethical and broader issues of AI in research output production. As a result, we recommend that further consideration will be needed as to its impact on the REF after the 2028 exercise.

We recognize our recommendations call for a significant shift in the REF process and that our role is purely advisory. But there is a significant change happening globally in research assessment because of the recognition of the unintended consequences of the traditional output weighted approaches. If the UK wishes to remain at the forefront of the global research effort as it should, it too must be willing to evolve its system of incentives. Traditionally the UK has been a leader not a follower in the issues of research assessment. It is thus important that the funding bodies should continue to engage internationally, for example, engaging with the Global Research Council and with the Coalition for Advancing Research Assessment (CoARA).

There remain emergent challenges which will need further consideration. There is work that needs to be done both to reduce cost and to build the incentives on the institutions by the greater weighting of university level strategy and activity. It is almost certain that further evolution of the REF will be needed after 2028. Any change in research assessment will have both intended and unintended consequences.

The funding bodies are to be congratulated on their ongoing willingness to review its approach to ensuring a high quality and adaptive research environment acknowledging the very many and contested views on research assessment from within the stakeholder communities. We thank the officials of the four funding bodies for their support of our work.

I thank you for the opportunity to be engaged along with the other members of the International Advisory Group on this important but complex and sometimes contentious matter.

Your sincerely

A handwritten signature in black ink, appearing to read 'Peter Gluckman', followed by a period.

Sir Peter Gluckman ONZ KNZM FRSNZ FMedSci FISC FRS

President International Science Council

Director, Koi Tū, the Centre for Informed Futures, University of Auckland
Chief Scientific Officer, Singapore Institute for Clinical Sciences.

Contents

Executive summary	6
Background.....	10
The IAG	13
The changing nature of academia.....	14
The issue of research assessment	15
<i>Global trends in research assessment</i>	16
<i>The intended and unintended consequences of research assessment</i>	18
The definitional problem	21
The problems of assessment	23
The submission.....	25
The changing purposes of the REF	28
<i>Inter and transdisciplinary research</i>	30
<i>Research policy and funding</i>	30
<i>Institutions and disciplines</i>	31
<i>Attributes, skills, and expertise</i>	31
Future challenges	31
Recommendations.....	36
Appendix A – IAG members.....	38
Appendix B – IAG meetings	39

Executive summary

The International Advisory Group (IAG) to the Future Research Assessment Programme (FRAP) comprised international experts in different aspects of research assessment from research production to research users. It was established in 2021 and met over 24 months to provide high level advice on the state and evolution of the UK's university research assessment process, the current version of which is termed the REF. The panel's terms of reference restricted its considerations to the general principles and did not consider operational details.

The panel met virtually and undertook its work in two phases. In 2021/2022 it focused on the general principles of research assessment and in 2023 it developed specific recommendations relevant to the future nature of the REF.

The panel concluded that while there are many views, both positive and negative, with the REF, the REF process has assisted in ensuring the standing of UK academic research globally.

The dual funding system which distributes > GBP 2bn as a result of the REF is a critical part of the UK higher education system and should be retained. However, the assessment process should continue to evolve reflecting both the changing nature of academia and evolving understandings of research assessment and to address a number of issues that have been identified.

There are changing views of research assessment globally and a diminishing focus on classic bibliometric approaches. The panel concludes that it is desirable that the UK follows this trend and reduces the weighting of outputs significantly.

Research excellence and impact, however they are defined and measured will never have the precise objectivity that some expect of it. This expectation is directly linked to the reality that the REF is being used to distribute significant funds and this drives behaviours, some of which are concerning, by institutions and academic units. Indeed, much of the criticism levelled at the REF comes from consequential institutional behaviours rather than governmental expectations. Our recommendations that shifting the burden away from a focus on the individual to the institution will hopefully assist in reducing these concerning aspects.

The purposes of the REF have grown and thus become confusing. The panel recommends these are simplified, condensed, and clarified with the main purpose being that of accountability and funds allocation so as to promote institutional behaviours that foster research excellence.

The weighting of traditional output assessment should be markedly reduced as this is a primarily measure of past individual or team-based activity and much of it has already been assessed in other ways. Impact assessments should give greater weight to 'normal' rather than 'extraordinary' impact. Assessment panels should seek to promote intellectual diversity and innovation and boundary activity that crosses traditional disciplines. Team based activity and the roles of the full university staffing

enterprise need to be considered. Transdisciplinary and interdisciplinary research needs to be particularly recognised and promoted.

Given that the block-grant funding goes to the institution for use at its own discretion far greater weighting must be given to the institutional environmental assessment, acknowledging that measures of the environment will need to be developed and agreed. We suspect that further evolution of the REF will be required after the 2028 round. Given the rapid emergence of AI-related techniques affecting both research production and outputs further work on this dimension will be needed before the next round of research assessment.

The panel notes with appreciation the quality of officials' assistance with its work and their deep knowledge of the issues that it needed to consider.

The panel makes 13 recommendations.

1. That the basis of the REF be retained and continue to be linked to a dual-support funding system research with the UK's higher education facilities.
2. That the purposes of the REF are clarified and simplified in the context of the changing nature of academic enquiry. These would be:
 - a. To continue to provide a source of institutional funding to promote research linked to past and expected performance,
 - b. To give the UK and devolved UK Government's and UK taxpayers confidence that the system is accountable for use of taxpayer funds and will sustain and promote UK's global standing,
 - c. Provide system wide and institutional data for performance enhancement.
3. That, if desirable, consider other ways of incentivising other activities beyond these core purposes with dedicated funds thus clarifying the link between REF and HEI performance and both strengthening and weakening (depending on the specific objective) the link between REF and funding.
4. Value the processes and environments that increase the probability of generating high quality and impactful research more than the past outcomes.
5. Recognise that the nature of research is evolving. In that context any REF must now recognise team-based research and the contribution of a broad scope of individuals who have in the past not been assessed. It must recognise the growing and essential contribution of under-represented communities.
6. Define, support and develop better processes to incentivise and support inter- and transdisciplinary research. The issues here extend beyond research assessment to both funders and employers.
7. The balance between outputs (contribution to knowledge and understanding), impact (engagement and impact), and environment (people, culture and

environment) needs to change. We recommend each should be weighted in the next round at 33%.

8. The definitions of outputs (contribution to knowledge and understanding), impact (engagement and impact), and environment (people, culture and environment) require revision with greater clarity of what is sought:
 - a. Redefine excellence/quality research in a more pluralistic manner
 - b. Reduce the system-wide focus on classic quantitative bibliometrics in assessing outputs
 - c. Break link to individuals
 - d. Seek ways to support distinctive profiles of units and true innovation in unit profiles
 - e. Impact must move from singular focus on exceptional impact to promote both exceptional and ordinary and sustained impact and stakeholder engagement
 - f. The research environment should be assessed at the institutional level and include measures to recognise and reward positive research cultures and environments
 - g. Research outputs should be assessed at the unit level not the individual level
 - h. Transdisciplinary and interdisciplinary outputs and impact are likely to involve multiple units and potentially institutions and the evaluation process must recognise and incentivise this
9. Recognise that diversity and plurality in the range of academic research in the UK is critical as much as current outcomes. Therefore, panel membership and criteria should reflect the evolving nature of research and the changing purposes of the REF as revised
10. UK funding bodies and HEIs should collaborate to review what data are collected and the rationale for its use. UK funding bodies, funders and institutions should look to reduce the time burden placed on academics and the overall cost of the REF system.
11. HEIs should be encouraged to reduce the burden associated with their internal use of the REF and giving it excessive emphasis (for example talking about 'REF-able' and 'non-REF-able' research or researchers). The REF should not be used as a substitute for staff performance or other within-institution activities or seen as a ranking tool.
12. UK funding bodies, funders and institutions should monitor and evaluate the impact of the REF on research culture and the UK's research and innovation landscape as a whole.
13. Note that it is likely that further evolution of the REF will be needed after 2028. Any change in research assessment will have both intended and unintended

consequences. There remain emergent challenges that this report has highlighted which will need further consideration.

Background

1. The nationwide assessment of research quality in the UK, known as the Research Selectivity Exercise (RSE) and Research Assessment Exercise (RAE) previously, and Research Excellence Framework (REF) in the last two cycles, started more than 30 years ago.
2. The first RSEs conducted in 1986 and 1989 were rather light-touch exercises. They were motivated mainly by the perceived need to reassure HM Treasury that funding provided for research was serving to maintain or advance the research base. These exercises were named as such because 'selectivity', aimed at allocating more resources to those who perform best, was seen as necessary given that the previously used 'equity principle' (that assumed all universities did research and needed institutional funding for it) was seen to favour traditional elite universities as well as not being transparent.
3. The RAE in 1992 saw the introduction of a more structured approach to the information requested from institutions (with restrictions to the number of submitted outputs and reductions to the number of disciplinary panels; this was further reduced over time), where they submitted a templated portfolio for each Unit of Assessment (UoA) containing data on staff, funding, students, publications and as well as strategic statements.
4. However, concerns over the costs and burden associated with the RAE exercises in 1996 and 2008, led to the development of a new exercise called REF. REF was designed with the goals of reducing the administrative burden on participating institutions, promote equality and diversity in research and avoid creating unhelpful behavioural incentives. The first REF was carried in 2014 and this exercise also saw the introduction of the assessment of impact, 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' through the submission of impact case studies.
5. The outcomes of REF, and exercises before it, have been used to inform the methodology of the formulae used by the four UK higher education funding bodies (hereon referred to as "funding bodies") (Research England, the Scottish Funding Council, the Higher Education Funding Council for Wales, and the Department for the Economy, Northern Ireland) to allocate institutional funding. It is important to note that the outcomes of the REF are used differently by the four funding bodies in the allocation of institutional funding. In addition to funding allocations, REF also provides accountability for public investment as well as providing significant benchmarking information that affect institutions' research reputation.

6. Significant effort goes into making submissions but success in REF leads to not only prestige but also to a highly sought and relatively stable, stream of income from government in the form of quality-related block grant research funds (over £2 billion across UK higher education institutions in 2022/23). However, REF has been described as a zero-sum game as increases in the block-grant budget are hard to come by, while the size of the research enterprise grows, and the evidence suggests that research quality inevitably improves in response to such incentives.
7. There is no doubt that REF and its predecessors have contributed to an improvement in the UK research base and the research performance of UK institutions locally and internationally, such that the UK is able to retain its world-leading reputation. Indeed, politicians (amongst others) might wonder whether this progress could be sustained without a research assessment process such as the REF. However, there has also been a cost arising from the complex and lengthy machinery that REF represents, both for research and for researchers.
8. Over the years, the competition between higher education institutions has sharpened and the focus on research performance has led to a systemic shift in culture as evidenced by centralisation of funds, departmental re-structures, appointment of research managers with specific research missions, formation of research strategy committees, etc. University managers value the REF's capacity to help them identify areas of research that their institutions might consider whether to expand or contract. It has, however, been argued that this reshaping has resulted in less diversity and more disciplinary conformity in the system. Issues also arise from attempts to 'game' the system (e.g., a REF-created transfer market and unfair hiring practices), potentially impeding research cultures.
9. Some have argued that in its original form, the RSE aimed at making funding allocations more transparent and selective by rewarding quality. However, the massification of higher education and transformation of polytechnics into universities in the early 1990s meant that over time, finite resources became further concentrated in areas of pre-existing excellence. On the other hand, there is recognition that REF has allowed identification of pockets of excellence and demonstrated UK universities' significant research capability across the board. There have been other criticisms levelled at REF, some justified and others not. For example, it could be argued that while participation in the REF is burdensome, some institutions have substantially and unnecessarily increased this burden through internal 'mock REF' exercises.

10. While there is much variation in perspective, some academics have expressed concerns about the potential impact of the REF on their work, whether directly or indirectly. Directly, the effort involved in reporting, managing, and monitoring research performance takes away the time used to carry out the research itself. Indirectly, the absence of comparable financial incentives for teaching has unwittingly linked academic success and prestige to research (to the detriment of teaching and learning as a valuable and important part of an academic's work). The REF has been criticised for its focus on a narrow set of research outcomes: publication and citation have become ends in themselves, disregarding the rich variety of outputs that can come from research or the huge rewards that can be obtained from collaborative work and diverse contributions or contributors.
11. There has been a trend to define research activity as REF-able or non-REF-able which may have inhibited some forms of needed scholarship and research activity – especially that which does not easily align with the unit of assessment. The REF has also been criticised for discouraging non-conventional, novel and applied research, thereby reducing the diversity of research ideas or methods.
12. The negative implications of REF are particularly felt by certain groups of academics such as early career researchers, especially where REF has been used as a performance management tool. It must be noted that there are misconceptions and 'myths' through which academics misinterpret REF rules, which will undoubtedly aggravate academics' feelings around feeling 'hard done by.'
13. Irrespective of how one views the REF, many in the sector understand that a system like the REF is needed for accountability of public money going into research and that there will never be a perfect assessment system. Many also have shown appreciation of the funders' willingness to work with the sector to address REF issues and the emphasis they place on keeping burden proportionate.

The IAG

14. The four UK higher education funding bodies established an International Advisory Group at the start of the Future Research Assessment Programme (FRAP) in May 2021 to assist and advise them in their development and evaluation of options for a future approach to national research assessment within the UK¹. The IAG was established with the purpose of providing independent and external perspectives and insights into research assessment principles and practices. The group was also tasked with acting as a 'sounding board' against which the thinking of the FRAP could be tested and the assumptions and scope of the review challenged where appropriate. It was always intended that the IAG would approach its work focused on high-level principles and make high-level recommendations to the FRAP Board on the most appropriate model(s) for future research assessment systems in the UK. As reflected in its Terms of Reference, it was never intended that the group would delve into the technicalities and finer detail of implementation.
15. The group was composed of individuals with a diverse range of expertise and experience of varied national contexts; this diversity of perspectives proved invaluable in the IAG's ability to challenge and offer sound advice. Members were sought from the following roles:
 - a. senior leaders of research funding organisations globally;
 - b. international experts in research evaluation;
 - c. non-academic research users.
16. The list of members is provided in Appendix A.
17. The group carried out its work in two phases. Phase one, conducted throughout 2021, consisted of six meetings and was about exploring principles and core big issues of research assessment (such as culture, excellence, and impact) and involved members bringing 'provocations' on these important topics for the group to discuss.
18. Phase two started in early 2022, when results from some of FRAP's commissioned pieces of review work as well as consultation outcomes and stakeholder engagement activities were starting to emerge, allowing the IAG (together with the FRAP Board) to reflect on what the emerging findings may mean for a future system. During Phase two, the Group met five times over 2022 and the first quarter of 2023 with an extraordinary meeting held in September 2022 for the team working on Harnessing the Metric Tide to update the IAG on their work. Appendix B contains further information on meeting dates including a content summary for each meeting.

¹ Terms of Reference for the IAG available at: <https://www.ukri.org/publications/future-research-assessment-programme-terms-of-reference/>

19. The group met via videoconference. Additionally, some of its business was conducted by correspondence between meetings. The meetings were attended by IAG members and various 'officials'². Attending Officials did so as observers or contributors to the meeting. The Associate Director for Research Environment at Research England represented the Board at all meetings. The secretariat was serviced by Research England on behalf of the four funding bodies. In between meetings, a short de-brief and agenda-setting meeting for the next IAG meeting was held between the Chair of the IAG and the Associate Director for Research Environment at Research England.

The changing nature of academia

20. While Universities have often been seen as very conservative organisations, the nature of academic research has been undergoing rapid change in a context where there has also been significant institutional evolution. The size and complexity of the operation of a modern university has brought with it much greater managerialism. Many academics would argue that this has changed their roles and reduced their flexibility in research and creativity. Added to that, the exit of the UK from the European research area has added increasing concerns about how the extraordinary high impact and high quality of the UK's academic research enterprise can be maintained. The recent Nurse review³ points to potential further changes in the structure of research funding. Notwithstanding a somewhat challenging environment, academic research and creative activity in the UK continues to show innovation and the UK remains at the forefront of the global research enterprise.

21. Over the past 70 years since Polyani and Merton both described the isolation of academic research from society^{4,5}, much has changed. Universities recognise their obligation to not only generate new knowledge but also to produce impactful and actionable knowledge. The 2014 shift in REF to include impact reflected this shift, which in a world with compounding challenges needs to continue to be incentivised.

22. However, change is afoot in how research is conducted in many ways and those changes need to be recognised in the evolution of research assessment. In some domains, research is increasingly done in teams. In many other domains, research increasingly involves framings and investigators spanning disciplinary boundaries and increasingly, stakeholders from beyond the university in a variety of modalities. The push for greater actionable knowledge

² The secretariat, the FRAP project team and FRAP Board members are referred to as 'the Officials' by the group.

³ <https://www.gov.uk/government/publications/research-development-and-innovation-organisational-landscape-an-independent-review>

⁴ M Polyani, The republic of science; its political and economic theory. Minerva 1 54-73 1962

⁵ RK Merton The sociology of science: theoretical and experimental observation University of Chicago Press 1962

is leading to the emergence of transdisciplinary approaches. These take time and require genuine non-academic stakeholder engagement. It needs to be appreciated that in transdisciplinary research, the research protocol is not defined a priori but emerges in the course of the dialogue with stakeholders. New funding tools may be needed in this area but universities themselves can do much to inhibit or foster such activity.

23. As team-based research and creativity grows, so does the importance of university members beyond the academic investigator; in assessing the research environment, how these individuals are recognised and supported becomes important. These may be technical professionals with deep expertise; they may be facilitators who engage with different publics in ways that make them true partners in research.
24. An overdue and critically welcomed feature of research has been the greater inclusion of minorities and women in all aspects and at all levels of the research enterprise. Actions such as the Athena-Swann Charter have assisted in this. Institutions and the research enterprise are better off for this. In assessing the research environment, the need to continue giving focus to such activities remains.
25. An area which needs greater reflection is that of the impact of large language models and generative Artificial Intelligence (AI). ChatGPT, while a primitive example of generative AI, might be used to write research outputs, although the technology is too new to assess definitive risks to established norms. This has the potential to change much about how research is conducted (with or without human involvement), how and where it is reported and how it is assessed. It is too early to know how the science system as a whole will put limits (if any) on the use of these technologies in the production of knowledge or at least its reporting but it has enormous potential to disrupt traditional forms of peer review and thus assessment of research outputs. The speed of these developments means that further work and review may be needed before the next research assessment round.

The issue of research assessment

26. Research assessment can serve many different purposes – for funding a proposal, for evaluating an individual, for assessment results for publication, and in each of these cases the questions to be asked are specific to the individual or the project. The REF has a fundamentally different purpose and thus it is not necessarily appropriate simply to apply the same processes. The primary purposes of nationwide assessment of research are generally related to either funding distribution (e.g., UK, New Zealand), providing accountability to the taxpayer, or to support the marketing of the Institution through rankings etc.

27. The origin of the UK research assessment exercise was to create a dual-support funding mechanism by which accountability for the research spend was increased and incentives were put in place to enhance the research performance. From the outset the challenge has been the difficulty in defining terms like research quality, excellence, and impact. Much of the burden of the REF comes from trying to achieve objectivity to the assessment of such terms which have inherent subjectivity. As discussed below, the objectives placed on the REF have grown explicitly and implicitly and while some were generated by government, others are generated by the higher education institutions themselves. The question that needs to be asked is whether the burden has grown larger than the value of the exercise.
28. Our view is that provided significant shifts are made in what is assessed, how it is assessed, and that the focus shifts from the individual to the institution, the REF will remain an important component of the UK funding mechanisms for higher education and it should continue. However, some change is needed.

Global trends in research assessment

29. Research assessment is a complex and nuanced process, and good practices vary depending on the specific context and goals of the assessment. However, there has been a decade of global dialogue on the need to reform research assessment⁶. The concerns and limitations of current practices have been highlighted, to improve the quality, validity, and fairness of research evaluation. In the following paragraphs we highlight some of the initiatives that have been developed or suggested to promote a responsible research culture and shift towards more qualitative and holistic approaches to research assessment. They provided a useful framing for the work of the IAP. However, it needs to be noted that the bulk of this work focuses on research assessment for individuals and research groups rather than as a tool for institutional core funding – this purpose as is discussed elsewhere in the report has implications for the assessment process.
- a. The San Francisco Declaration on Research Assessment (DORA) is a set of recommendations developed in 2012 by a group of international scholars and publishers that aims to improve research assessment practices. The declaration encourages a shift away from using journal-based metrics (such as impact factor and h-index) as a primary means of evaluating research, which can create biases and lead to the undervaluation of research in certain fields or regions. Instead, it advocates for a more holistic and qualitative approach to assessing research quality and impact. The declaration calls for the assessment of research to be based on multiple criteria such as the rigour of research methods, the quality of data, the significance of the findings and the potential for practical application and impact.

⁶ For a recent review see *The future of research evaluation: a synthesis of current debates and developments: (2023) DOI: 10.24948/2023.06*

- b. The Leiden Manifesto for Research Metrics is a set of principles and guidelines developed in 2015 in response to concerns about the misuse of research metrics. The ten principles of the Leiden Manifesto call for the responsible use of metrics in research evaluation, emphasising the need for context-specific evaluation of research, rather than relying on a single metric. The principles recognise that research metrics have limitations and that they should be used in conjunction with other evaluation methods, such as peer review and expert judgment. It also emphasises the need for transparency and the importance of communicating the limitations and appropriate use of research metrics.
- c. Research on Research Institute (RoRI) is a global research initiative which brings together a network of researchers, policymakers, funders, and other stakeholders to identify key challenges in research and develop evidence-based solutions. Its research agenda includes topics such as research integrity, research culture, research metrics and research evaluation. It launched a working paper on 'The changing role of funders in responsible research assessment: progress, obstacles and the way ahead'⁷ setting out the state of play regarding RRA, exploring what RRA means and outlining existing principles and frameworks in the space.
- d. In July 2020, Science Europe published a position statement and set of seven recommendations to guide evaluation and assessment processes, for both research funding and research performing organisations. Namely: i) Research assessment processes must be clear and transparent; ii) Research organisations should monitor and regularly evaluate the robustness of their; iii) assessment processes, and share best practices to foster mutual learning; iv) Research organisations should publicly demonstrate and continually evaluate how they address bias, discrimination, and unfair treatment in assessment processes; v) Research organisations should streamline assessment processes to reduce the burden on reviewers and applicants; vi) Research assessments should focus on the substance and content of applications; vii) Research organisations should consider implementing novel assessment techniques. Their recent paper, 'Recognising What We Value'^{8,9}, provides practical recommendations and good practice examples detailing how research organisations can continuously improve the way they assess research and researchers.¹⁰

30. Momentum has gathered around the need for collective reform of research assessment systems, as seen through the establishment in 2022 of the Coalition for

7

https://rori.figshare.com/articles/report/The_changing_role_of_funders_in_responsible_research_assessment_progress_obstacles_and_the_way_ahead/13227914

⁸ <https://scienceeurope.org/our-resources/recognising-what-we-value-recommendations-on-recognition-systems/>

⁹ <https://www.universitiesuk.ac.uk/topics/research-and-innovation/uk-forum-responsible-research-metrics>

https://www.youtube.com/channel/UCeuLK_JPExentZ1XA8bGDWw?app=desktop

¹⁰

https://www.globalresearchcouncil.org/fileadmin/documents/GRC_Publications/GRC_RRA_Conference_Summary_Report.pdfhttps://www.globalresearchcouncil.org/fileadmin/documents/PDF_Links/Responsible_Research_Assessment_Call_to_Action.pdfhttps://www.globalresearchcouncil.org/fileadmin/documents/GRC_Publications/GRC_RRA_Conference_Summary_Report.pdfhttps://www.globalresearchcouncil.org/fileadmin/documents/PDF_Links/Responsible_Research_Assessment_Call_to_Action.pdf

Advancing Research Assessment (CoARA11)The CoARA Agreement¹², developed by the European University Association (EUA), European Commission and Science Europe, proposed four concrete core commitments to implement a global reform of research assessment. These are: i) Recognise the diversity of contributions to, and careers in, research, in accordance with the needs and the nature of the research; ii) Base research assessment primarily on qualitative evaluation for which peer-review is central, supported by responsible use of quantitative indicators; iii) Abandon the inappropriate uses in research assessment of journal and publication- based metrics, in particular the inappropriate uses of journal impact factor (JIF) and h-index ; iv) Avoid the use of rankings of research organisations in research assessment.

31. Both, DORA and CoARA are complementary initiatives with the same goals. CoARA provides a framework and timeline to operationalise changes¹³. The developments above have created the basis for a Global Research Council Statement of Principles, currently in preparation, which will be adopted during the GRC 11th Annual Meeting¹⁴. The Statement of Principles on Recognition and Rewarding Researchers will emphasise the need to develop new ways of assessing research and innovation at a global scale, adhering to high standards of quality regarding the research and research processes, however opening up to the recognition of new types of outputs (software, policy, societal dialogue, science communication, etc.)

The intended and unintended consequences of research assessment

32. Clearly, any research assessment exercise has the primary purpose of quality assurance, be it for a research grant, publication, or institutional evaluation. But in the case of the REF the potential for its misuse and for a wide range of unintended consequences has been raised repeatedly.
33. The UK maintains a remarkably high reputation for its research quality across multiple domains of endeavour despite the difficulties of Brexit and ongoing uncertainties of access to the European Research Area. We suspect the REF has contributed to this reputational positioning.
34. In the case of REF and some other institutional research assessment exercises (e.g., New Zealand) the exercise is directly linked to funding. This singular decision both changes the magnitude of the exercise and its significance from being that of individual, unit, or institutional reputation, to being a competition for funding. As in any incentive system this drives both intended and unintended behaviours.

¹¹ <https://coara.eu/>

¹² https://coara.eu/app/uploads/2022/09/2022_07_19_rra_agreement_final.pdf

¹³

https://rori.figshare.com/articles/report/The_changing_role_of_funders_in_responsible_research_assessment_progress_obstacles_and_the_way_ahead/13227914

https://rori.figshare.com/articles/report/The_changing_role_of_funders_in_responsible_research_assessment_progress_obstacles_and_the_way_ahead/13227914

¹⁴ <https://globalresearchcouncil.org/news/2023-annual-meeting/>

35. A key question must be whether it continues to serve its intended purposes. If the core purposes of the REF are those of distributing funding on a research quality criterion and providing accountability of publicly funded research to the funding administration, then these must drive how its future is considered. That the REF has been sustained over multiple cycles suggests that it has the overall confidence of the research sector, much of the academic community, and government and its agencies.
36. The question of whether the REF has improved research quality (which would be its underlying objective from the academic's perspective), is very difficult to assess as the definition of research quality is, as the rest of this report suggests open to considerable debate. There is no easy way to assess the counterfactual of no research assessment exercise. Essentially all nations with a strong research, development and innovations focus either have a formal evaluation system like the REF or effectively have an informal one through university ranking systems.
37. The diversity of funding created by the dual-support funding approach should not be given up trivially as it allows institutions some discretion and innovation in how they evolve.
38. The 2014 decision to include impact as a variable has clearly had impact (pun intended) on the higher education sector as they have made great progress in demonstrating how the research that they support has multiple pathways to impact. The consequence of this singular development was a changed attitude of universities and their staff to their impact on society, perhaps somewhat reducing the walls around ivory towers. This changed nature of the relationship between scholarship and society is critical if universities are to meet their obligations to the taxpayer. The various dimensions of benefit to society via public good, the policy sector and business must all be reflected in measures of impact and quality. But in turn such changes have been progressive over the past few decades globally.
39. There is evidence that over time, higher education institutions have used the system to weigh their strengths and weaknesses and rebalance their portfolios accordingly. This is important, as the diversity of their nature and capacities is one of the sector's strengths. The key issue is for the REF to thus reflect that diversity. The recent book *The Quantified Scholar*¹⁵ raises the issue as to whether the REF has diminished disciplinary diversity across institutions to meet some perception of the ideal unit as perceived by an assessment panel. This would be an unfortunate consequence and we recommend processes are put in place to promote innovation, diversity, and distinctiveness rather than homogeneity at the unit level. Minority and boundary disciplines or subdisciplines which have research conducted only in a small number of

¹⁵ J P Pardo-Guerra (2022) *The Quantified Scholar*; How research evaluations transformed the British social sciences. Columbia University Press

institutions may be at a disadvantage; assessment panels must be designed in a way to avoid this risk.

40. However, the unintended consequences of research assessment are not trivial. First is that of cost and burden on institutions and staff. What is less clear however is how much of that is driven by the need to meet the REF returns and how much it is driven by the utility of the measures to university leaders for other purposes. This has dangers especially as an incentive scheme developed for one set of purposes may not be suitable for co-option for other purposes. A case in point may be that described in Nature¹⁶. The issue at hand is that of whether bibliometric assessments fairly reflect how research is done and its broader impact. Clearly, they do not, and this is one reason we recommend a reduction in the weighting of output measures.
41. A major issue is that the REF brings an inappropriate focus on individuals when research is often (at least in the natural and health sciences) a team game, incorporating the efforts of research professionals and support staff or students. Any progression in the REF process must reduce or avoid a focus on the individual and focus solely on the unit and the institution. It is unfortunate but appears unavoidable, that the REF system is somewhat disconnected from the other core attribute of a university, teaching, but this adds much complexity and undermines the core purpose of the REF. Institutional leaders must look at the unit in an integrated manner evaluating its research, teaching, and engagement. It would be unfortunate if they focus on the research component alone because of its fiscal and reputational consequences.
42. By recommending a major shift in the weighting from research outputs to impact and engagement, and people, culture, and environment, we hope that it shifts the focus from the individual to the unit and institutional leadership. After all the funding does not go to the individual academic but to the institution as a whole and it should be accountable for how it is used to produce a stronger university across all dimensions expected of it. Part of that, as is discussed elsewhere, is staff development, promoting equity and diversity, and of relevance, the development of new domains of enquiry particularly through inter- and transdisciplinary activity.
43. Any incentive scheme will inevitably face some gaming. That is why designing the scheme to be clear on the behaviours desired and discourage negative consequences is so important.
44. Within institutions, an industry of impact staff has built up. It is important that their primary role is to encourage impactful behaviours rather than to narrowly focus on the REF. In some institutions, research assessment has become a topic of academic enquiry and that is highly desirable as clearly there remains much to learn.

¹⁶ Nature 595, 471-472 (2021) doi: <https://doi.org/10.1038/d41586-021-01991-z>

The definitional problem

45. Mechanisms for assessing research performance and articulating the impact of research whether in the academic realm or of society are confounded by the desire to be objective: but the terms used such as "excellence" and "impact" convey different meanings to different audiences. For example, traditional assessments consider academic impact as the intellectual contribution to a specific field of study through publication but do not necessarily look at the impact on society or the contribution to industrial applications which may be harder to define. To further complicate matters, the nature of modern research has evolved across and between traditional disciplinary research boundaries with the emergence of new disciplines and the rising importance of collaborative and inter/transdisciplinary research to address contemporary research questions. To that end, there are no uniform definitions, no set of objective outcomes, nor measures in research that enable research assessment exercises to recognise diversity in research roles, careers, practices, and activities, or the varied and broad contributions by researchers outside of the traditional lens, or in other academic endeavours (e.g., teaching, patient care, public and policy engagement, academic citizenship). On the contrary, established assessment exercises attempt to apply normative definitions to criteria such as "quality, excellence, impact, and environment" and ignore the contribution that diversity across disciplines brings to the overall valuation.
46. A major aim of research assessments is to ensure that research meets a minimum standard of quality. However, neither the definition of, nor the requirements for, research quality are standardised across research disciplines and institutions. On the contrary, narrow criteria and indicators of research quality that include individual, and publication-based metrics as proxies for research quality fall short of recognising diverse disciplines, research practices and contributions outside of the standard lens. The ability to gauge the quality of research must therefore be carried out through transparent research methods and processes that enable its review, critique, validation, reuse, and extension to advance knowledge and its potential impact.
47. The notion of excellence permeates throughout academic research and research-related activities at individual, institutional, and national levels. It is a pervasive term used in the evaluation of research to signify a standard of quality. However, there is no consensus, standard definition, or reliable measure of excellence. The evaluation of excellence involves highly subjective value judgements, and is commonly associated with quantitative metrics (e.g., journal impact factors, H-index scores, and university rankings). The vagueness surrounding notions of excellence has performative effects, and promotes certain disciplinary norms, while discounting others. Therefore, the notion of excellence is highly contested; there are concerns around its opaqueness which undermines trust with the research community, and concerns that the notion of excellence perpetuates hyper-competition and

inequities in the allocation of research funding. Alternative and more inclusive definitions of excellence that promote pluralising or diversifying excellence include a range of perceptions of performance that consider local context and practices.

48. The REF defines impact 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." However, the transition from research to impact is not always a linear pathway as research typically builds upon a body of knowledge and research that preceded it. Sivertsen and Meijer¹⁷ distinguish between normal and extraordinary impact. They posit that the impact of research does not occur primarily as unexpected and extraordinary incidents of breakthroughs (extraordinary impact). Rather, they suggest that impact results from normal everyday interactions among organisations, individuals, and fields of research that occur over short-to-long timescales. Requiring researchers to demonstrate a direct or causal link between their research and impact can potentially stifle creativity and innovation and potentially ignores the contributions to, and nature of, impact across disciplines and research types. It is also complicated by issues of attribution with often multiple contributions from multiple research activities over time adding together to create impact. Therefore, research assessment exercises should seek to highlight all forms of impact rather than only extraordinary impact and the extent to which organisational- and individual-level activities and responsibilities support these interactions.
49. The research environment - which includes the strategies, infrastructures, facilities processes, and management systems - is inextricably linked to research culture (the behaviours, values, expectations, attitudes, and norms of research communities)¹⁸. However, there is no clear consensus on what institutional characteristics or capabilities constitute a thriving research environment or culture and therefore, there are no standard criteria for assessment. Instead, it can come down to a subjective assessment; for example, REF2014 asked the panels "Does this place look like a great place for work where senior and junior researchers will move the discipline further?"¹⁹ with attempts to look at how the institutions provided evidence of strong, positive environments where research could thrive.

¹⁷ Gunnar Sivertsen, Ingeborg Meijer, Normal versus extraordinary societal impact: how to understand, evaluate, and improve research activities in their relations to society?, *Research Evaluation*, Volume 29, Issue 1, January 2020, Pages 66–70, <https://doi.org/10.1093/reseval/rvz032>

¹⁸ <https://royalsociety.org/topics-policy/projects/research-culture/>

¹⁹ Pidd M., Broadbent J. (2015) 'Business and Management Studies in the 2014 Research Excellence Framework', *British Journal of Management*, 26: 569–81.

50. Therefore, creating a better understanding of what high quality research environments should look like seems imperative. There are pervading views that a thriving research environment, in addition to displaying adequate funding and high graduation rates, should embed principles of equity, diversity and inclusion, research integrity, open science, and inter/transdisciplinary working.

The problems of assessment

51. Peer review is widely considered to be an objective way to assess the quality, relevance, and outcomes of research, and a transparent means by which the research community feel they can trust funding allocations are made fairly. However, there are known limitations in peer review, for example:

- a. Peer review panels generally can identify weak from good proposals but struggle to reach consensus between proposals of good (or excellent) quality from proposals of exceptional (or outstanding) quality.
- b. There is mixed evidence on the effectiveness of peer review panels to predict future productivity in the form of scholarly publications, citations, or follow-on funding.
- c. Wide inter-reviewer and inter-panel variation is observed in the assessment of individual research proposals leading to divergent outcomes.
- d. Social effects are observed in the dynamics of panels - such as dominant personalities, assignment of lead reviewers, and the deliberation process – which affect scoring decisions.
- e. Certain groups, often those that are under-represented or marginalised, are strongly disadvantaged in peer reviewed assessment processes.
- f. There is a tendency for peer review panels to favour shorter-term, lower-risk projects over longer-term, higher-risk innovative research.
- g. Interdisciplinary contributions and 'non-conventional' outputs (e.g., data, tools, and technologies) are often not adequately credited or recognised.

52. Despite their known limitations, quantitative metrics, such as counts of publications, citations, h-index, and league tables, are often used to assess research performance, quality, and impact. More than a decade after the Leiden Manifesto²⁰ and the San Francisco Declaration on Research Assessment²¹ were launched to abandon journal impact factors (JIFs) and promote the responsible use of metrics, there remains much anecdotal evidence the use of narrow quantitative measures include²²: "distorting research incentives, creating unsustainable pressures on researchers, and

²⁰ <http://www.leidenmanifesto.org/>

²¹ <https://sfdora.org/>

²² Curry, S., de Rijcke, S., Hatch, A., Pillay, D., van der Weijden, I. and Wilsdon, J. (2020) The changing role of funders in responsible research assessment: progress, obstacles & the way ahead. RoRI Working Paper No. 3., November 2020. <https://doi.org/10.6084/m9.figshare.13227914>

exacerbating problems with research integrity and reproducibility; [reducing the] diversity of research missions and purposes... to adopt similar strategic priorities, or to focus on lower-risk, incremental work; [propagating] systemic biases against those who do not meet—or choose not to prioritise—narrow criteria and indicators of quality or impact, or to conform to particular career pathways; and divert[ing] policy and managerial attention towards things that can be measured, at the expense of less tangible or quantifiable qualities, impacts, assets and values – a trend exacerbated by the rise of flawed university league tables." The reform of research assessment must include a more responsible use of quantitative metrics and the adoption of qualitative methods to capture a broader range of diverse outcomes, qualities, and impacts of research^{23,24}.

53. While the processes and mechanisms for allocating research funding are similar among many national systems, institutions, and programmes, the factors that contribute to funding outcomes are not as well understood, nor the extent to which ways of allocating research funding become inefficient when award rates are low, or funding is limited. Furthermore, there is considerable concern that low award rates for research funding promote hyper-competition and distort research practice, since it may incentivise a priority for short-term gains, such as publishing in high-impact journals, rather than pursuing long-term research goals. Such "publish or perish" practices, can damage the integrity and rigour of scientific research. They can also create toxic research environments and cultures. It is therefore essential to consider how funding allocation processes can lead to different outcomes, to ensure that they promote an inclusive and sustainable research culture.
54. Research assessment processes need to be structured in a way to optimise decision-making and mitigate the factors that impair robust decisions. Further, careful attention should be given to panel membership to include members that hold diverse world views, cultural contexts, and positions on a problem and are enabled and supported by robust processes to work through opposing views to a conclusion.
55. There is considerable literature on the burden of peer review. The operating costs of research assessment have been estimated through work commissioned by the Funding Bodies²⁵. The full cost not only includes the administrative resources required to support the full research assessment process, but the cost to those engaged in the assessment process. Fatigue is rising among researchers asked to contribute to numerous at the expense of less tangible or quantifiable contributions or impacts. Some of this fatigue could be counteracted by measures to make peer review less burdensome.

²³ Science Europe. (2022) Agreement on Reforming Research Assessment.

<https://www.scienceeurope.org/news/rra-agreement-final/>

²⁴ Curry, S., Gadd, E and Wilsdon, J. (2022). Harnessing the Metric Tide: indicators, infrastructures & priorities for UK responsible research assessment. <https://doi.org/10.6084/m9.figshare.21701624.v2>

²⁵ See the Review of REF costs and benefits report at <https://www.jisc.ac.uk/future-research-assessment-programme/evaluation-activities> for details

56. There is growing interest in how AI might be used to support various research assessment activities – from automation of process tasks, to analysing quantitative data to inform decision-making, to mining text for thematic trends. There have been attempts to simulate human judgment using AI to identify features of research assessment that can be used to predict certain funding outcomes. Where machine-learning models were able to "predict" funding outcomes, predictors were factors not related to research impact, but to narrow quantitative metrics (e.g., journal impact factors, H-index scores, and university rankings). Significant challenges remain before machines can be reliably deployed to support human decision-making in the allocation of funding²⁶. While the case for AI to reduce the burden for menial tasks is growing, it remains to be seen whether the newest advances in AI (e.g., ChatGPT) will supplant peer review to assess the novelty, quality, or impact of research
57. The Funding Bodies' Cybermetrics and Research Evaluation Group at the University of Wolverhampton explored the use of AI in REF. They delivered their recommendations in December 2022. They concluded that AI is not currently able to replace peer review although an AI pilot could be launched parallel to peer review in the next REF to learn more about its potential. AI was seen to recapitulate existing biases within peer review. It is not expected that AI will ever have high accuracy of scoring since it cannot replicate the lifetime expertise of reviewers and it includes the use of metrics such as journal impact factor and is thus not compatible with the responsible use of metrics in research assessment. It is noted that these recommendations are based on the use of AI to assess research outputs, not research impact.

The submission

58. In the most recent REF 2021 exercise, participating HEIs were required to make a submission comprising a complete set of data about staff, research outputs, the impact of their research, and the research environment:
- a. Outputs are defined as the outputs of research, scholarship and creative endeavours including journal articles, books, conference proceedings, film, choreography, exhibits etc. Submission of outputs includes evidence of scholarly attention including citations, critical acclaim, and awards. Outputs counted for 60% in REF 2021.
 - b. Impacts are the provable effects (hopefully positive) of research in the "real world" including impacts on society, health, economy, environment, and culture. Impacts are described using structured case studies which must include evidence of the impact derived from testimonials, citations in policy documents or reports or minutes of community boards. Impact counted for 25% in REF 2021.

²⁶ <https://repository.jisc.ac.uk/8980/1/can-ref-output-quality-scores-be-assigned-by-ai-experimental-evidence.pdf>

- c. The Environment submission includes sections on; context and structure, research and impact strategy; people (including equity, diversity and inclusion as well as early career researchers); research income, infrastructure and facilities; collaboration and contribution to the research base, economy and society. The environment section counted for 15% in REF2021.
59. Our discussions on the burden of the REF considered the reality that many outputs have already been subject to peer review for journal articles, editing for books and critical acclaim for creative outputs. The IAG debated the relative balance between the three main elements, recognising that the weight for outputs should be lower since they are already subject to assessment. IAG debated on the merit of impact case studies (see below) and the opportunity for institutions to describe their commitments to impact activities in the environment section.
60. Research on REF 2014²⁷ concluded that the case study is an appropriate instrument for impact assessment. Although impact is by far the most expensive element of the REF the IAG recognised that there was no clear alternative. Nonetheless, the IAG recognised a number of challenges.
61. There were practical problems for academics, such as interpretations of guidelines, examples, FAQs etc. where there was some ambiguity. The reality of so many different pathways to impact was that it was difficult to provide clarity for all cases.
62. The guidelines for impact case studies are at best generic. Officials cannot provide specific advice for impact case studies. This results in impact officers and researchers struggling to make the "perfect" case study by spending increasing effort for decreasing returns. Case study authors strive for excellence where very good might be sufficient yet the competitive nature of the REF drives unhealthy behaviours. The extensive FAQs illustrate the degree of anxiety for case study authors.
63. The same research on the REF 2014 demonstrated that there were 3,709 unique pathways to impact in the 6,679 submitted case studies. This diversity of approaches to impact means there are no templated approaches to impact. Each reviewer is presented with a different case study with different sources of evidence and must approach each with fresh eyes. In addition, case study authors must start from scratch with each case study creating a burden on all elements of the case study production and review driving stress and expense in the system.

²⁷ Grant, J. and Hinrichs, S. (2015) The nature, scale and beneficiaries of research impact: An initial analysis of the Research Excellence Framework (REF) 2014 impact case studies.
[https://kclpure.kcl.ac.uk/portal/en/publications/the-nature-scale-and-beneficiaries-of-research-impact\(7ae869ae-6f1b-45b7-838e-48a270a08eea\).html](https://kclpure.kcl.ac.uk/portal/en/publications/the-nature-scale-and-beneficiaries-of-research-impact(7ae869ae-6f1b-45b7-838e-48a270a08eea).html)

64. Preparation of impact case studies can have issues with attribution²⁸. They are required to assess the connection between research outputs, their use in government, community or industry settings and the impact on the public by the ensuing public policies, social services and new products respectively. Most research is used conceptually not in an instrumental, linear fashion²⁹ so attribution remains a concern for case study authors and reviewers.
65. Some research activities take many years to produce broader impacts³⁰. In that time a researcher could have switched institutions, staff in partner/user institutions will have moved on to new roles or new employers creating barriers for the collection of testimonials and evidence of impact. This drives effort and thus expense. Conflicts on "ownership" of impact between institutions who have employed the researchers are possible.
66. Most "wicked problems" require evidence from different disciplines. For example, generating an impact on homelessness may involve research from poverty studies, criminal justice, child protection, mental health, LGBTQ+, Indigenous, public health and big data. However, the Unit of Assessment structure may drive disciplinary homogenisation, not interdisciplinary diversity and there may be challenges in linking impact case studies to diverse research outputs
67. As described by Silversten and Meijer³¹, "Societal impact of research does not occur primarily as unexpected extraordinary incidents of particularly useful breakthroughs in science. It is more often a result of normal everyday interactions between organisations that need to create, exchange, and make use of new knowledge to further their goals." With an emphasis on reach and significance of research impact, and rewarding both with funding (allocation), the REF Impact criteria encourage institutions to seek examples of extraordinary impacts while the policy directives to encourage research with an impact on society (accountability) can be served by supporting ordinary impacts especially when place-based universities are working in local contexts. This misalignment creates tension between institutions and policy directives particularly for researchers and impact officers seeking to draft impact case studies to meet multiple expectations. This leads the IAG to consider how an impact case study can be good enough to describe ordinary impacts without having to seek evidence for extraordinary impact by investing increasing effort for decreasing returns. How good is good enough for an impact case study?

²⁸ <https://www.thecatalyst.org.uk/resource-articles/measure-attribution-impact-evaluations>

²⁹ <https://kappanonline.org/new-day-education-research-practice-farley-ripple/>

³⁰ Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. *J R Soc Med*. 2011 Dec;104(12):510-20. doi: 10.1258/jrsm.2011.110180.

³¹ Sivertsen, G. and Meijer, I. (2020) Normal versus extraordinary societal impact: how to understand, evaluate, and improve research activities in their relations to society?, *Research Evaluation*, 29(1): 66-70, <https://doi.org/10.1093/reseval/rvz032>

The changing purposes of the REF

68. There are six potential purposes of research assessment:

- a. **Analysis:** To understand why, how and whether research is effective, and how it can be better supported.
- b. **Advocacy:** To demonstrate the benefits of supporting research and enhance the understanding of research and its processes among policymakers and the public.
- c. **Allocation:** To determine how to distribute funding across the research system.
- d. **Accountability:** To evidence that money and other resources have been used efficiently and effectively, and to hold stakeholders to account.
- e. **Acclaim:** To compare and recognise the value of higher education institutions and the research conducted within them.
- f. **Adaptation:** To steer change in organisational structures, behaviours and cultures, and research activities and priorities.

69. Different impact stakeholders will have different purposes for undertaking impact assessment that may not align with the primary purposes of the UK funding bodies REF: allocation and accountability. Institutions and units want acclaim and advocacy for more funding. Researchers and students may wish to learn from impact efforts (analysis) ultimately adapting to get better outcomes. Funders will want to be accountable to the government. The lack of alignment creates gaps in motivation among different impact stakeholders.

70. Over time the UK research assessment exercise has accumulated more purposes. From its original intent to provide accountability and to inform the allocation of funding, it has informally and formally gathered additional purpose some of which have influenced its design and others have affected the behaviours of the institutions and academics. The Stern review (2016)³² highlighted six objectives of the REF:

- a. Inform the allocation of funding
- b. As an essential tool of accountability for public investment in research in higher educational facilities and to support the case for additional funding
- c. To provide an evidence base for national decisions on research priorities
- d. To create a strong performance incentive for universities and for individual academics
- e. To create a tool for universities to inform their own decisions on resource allocation
- f. To provide a periodically updated reputational benchmark

³² <https://www.gov.uk/government/publications/research-excellence-framework-review>

71. While not explicit, the REF has appeared to have evolved another implied objective: discussion on the use of the REF to address other forms of culture change including the promotion of transdisciplinary research and address issues of diversity and career development in the workforce were heralded by Stern and have been prominent in discussions since then.
72. When there are multiple and not necessarily always coherent objectives and the priority of objectives varies when viewed by government, institution, and research team, confusion can arise and there can be little doubt this multiplicity of objectives has added to the burden of the REF. It is important to distinguish between those objectives that meet the needs of the funding bodies, from those that meet the interests of institutions. It is this plethora of objectives that adds to the cacophony of views about the REF – depending on who's perspective one looks at it. We believe these should be simplified, to focus on the REF's core roles of accountability and resource allocation. It may be, if there is a policy desire to promote additional objectives, that alternate mechanisms need to be found (e.g., discrete funding incentives for staff diversity).
73. One worrying trend is the REF can be used as a proxy for staff and unit management. The REF should not be seen as an individual performance metric and institutions should be required to have separate staff management systems. Further, the diverse scope of research needs to be encouraged and the drivers that are put on academics regarding what is "REF-able" are concerning. For this and other reasons we recommend a greater focus on the Institution and unit and a move away from a focus on the individual.
74. The related reason is that the funding goes to the university to use prospectively as it sees fit. We note the recent Nurse Review³³ recommends that the use of the funds should be transparent – however given that the dual-support funding system creates a bulk fund for institutions, while it is theoretically desirable and might build internal trust within institutions, it is not practical. While the funding goes to the university to use prospectively, the REF to date is largely based on retrospective performance by individual academics. This is an inherent mismatch – this is one reason why we favour giving greater weight to the performance of the institution as whole through the environment measure – it is the University that should as a whole be accountable for the decisions it makes as to how the research environment is developed.
75. A further reason for focusing on the University as a whole, rather than individual academics is the changing nature of academia and academic research that requires leadership and responsiveness by the institution as a whole. Much research is now team based and involves professional staff with distinct skills in areas ranging from technical to facilitation. Teams must be allowed to grow and evolve, and academic staff are but part of that team. Diversity matters – increasingly it is clear that past impediment to women and minorities are being

³³ <https://www.gov.uk/government/publications/research-development-and-innovation-organisational-landscape-an-independent-review>

addressed and institutions have a moral responsibility to continue efforts in that direction, with or without fiscal incentives. But diversity in knowledge generation requires breaking disciplinary silos, and the new tools of transdisciplinarity require institutions to look to their own structures. Traditional bibliometrics, siloed departments and internal budgetary rules all inhibit the development of actionable knowledge that can have impact in multiple ways. These are issues that only the University as a whole can address: the measures they choose to take can be reflected into an environmental assessment.

Inter and transdisciplinary research

76. Inter- and transdisciplinary research is of value in and of itself but might just provide the radical solutions for the monumental and 'wicked' global challenges which we face as a society. The problems of today are complex and multifaceted. Developing solutions demands a variety of disciplinary and practitioner approaches and perspectives and deep and meaningful collaboration.³⁴
77. There are enduring challenges for those engaged in inter- and transdisciplinary research, from the policy and funding landscape to institutional structures, and the need for personal and interpersonal skills opportunities across disciplinary and sectoral boundaries. The possibilities are ground-breaking: climate scientists working with artists; computer engineers working with philosophers; neuroscientists collaborating with dancers; and architects working with non-governmental organisation.
78. The main preconditions for successful inter- and transdisciplinary research can be grouped under three headings. The REF can shape each one, especially 1 and 2.

Research policy and funding

- a. The research policy and funding landscape matters. Long-term commitment to supporting inter- and transdisciplinary research, and to supporting capacity building for it, are needed.
- b. Challenge-led research questions create a good opportunity for collaboration, but much collaboration starts from the bottom up, from curiosity, and from supporting basic research.
- c. How funding calls are designed, and proposals evaluated, is important. Co-design with, for example, arts, humanities and social sciences scholars and experts can improve how problems are framed by funders, and inter- and transdisciplinary expertise is also needed in designing

³⁴ <https://www.shapeidtoolkit.eu> and <https://www.leru.org/publications/implementing-interdisciplinarity-in-research-intensive-universities-good-practices-and-challenges>

calls and evaluating proposals. With this in mind, a range of funding instruments are needed to build capacity and support more ambitious large-scale collaboration. Language of funding calls should be more open and inclusive, inviting a range of perspectives in addressing the topic and explicitly welcoming a broad range of contributions from all disciplines and stakeholders.

- d. Inter- and transdisciplinary research take time and trust, and the arts, humanities and social science disciplines in particular require support to build capacity. Seed funding is needed to build networks, consortia, and relationships, including with industry and societal stakeholders, laying the foundation for larger-scale collaboration.

Institutions and disciplines

- e. Higher education institutions and their disciplinary structures have a huge influence on researcher education, training, and careers. Inter- and transdisciplinary research is often seen as risky for early career researchers because of the lack of viable career pathways; culture change takes time. Policymakers and funders can support the development of inter- and transdisciplinary education and research in higher education institutions to build capacity from undergraduate to postdoctoral and faculty level, training in "meta-skills" and developmental support for those in institutional governance roles.

Attributes, skills, and expertise

- f. Interdisciplinary teams require a range of disciplinary expertise, integration expertise (people who can bridge the disciplinary knowledge) and also the involvement of relevant societal and enterprise stakeholders, and respect for their knowledge and experience.

Future challenges

- 79. The research ecosystem is in a continual state of flux with new expectations, new technologies and new policies challenging existing scholarly norms.
- 80. The 600-year-old model of a university that sits apart from and observes society is changing. There is increasing focus on the university as an engine for educating citizens and undertaking research that can address socioeconomic challenges. This new social contract is emerging as international rankings increasingly dominate all aspects of university business. New technologies are enabling new forms of research dissemination such as pre-prints (with or without peer review), dissemination on social media and a more diverse forms of research communication such as infographics. New policy frameworks are

driving universities to adopt equality, diversity, and inclusion (EDI), open science, open data, and research data management.

81. Universities worldwide are vying for position on various global rankings such as Times Higher Education (THE) and QS rankings and nationally universities are ranked on a variety of indicators including research. The REF contributes to competition for rankings in the UK. The IAG heard many depositions and presentations on rankings and responsible use of research metrics. Harnessing the Metric Tide³⁵ makes ten recommendations including avoiding "metrics only" assessment, reforming and simplifying REF over two cycles and cautioning the use of league tables. Harnessing the Metric Tide also recommends re-imagining and providing greater weight to the Environment section of the REF. This creates opportunities for institutions to describe not only infrastructure and services supporting research outputs and impact but how institutions support equity, equality, diversity, and inclusion including early career researchers in the research enterprise. The IAG is also mindful of global initiatives like the Declaration on Research Assessment (DORA). DORA recognises the need to improve the ways in which the outputs and impact of scholarly research are evaluated³⁶.

A summary of the arguments: Revolution or evolution?

82. There is growing concern about the cost and burden of assessment exercises. There are factors that contribute to driving cost: the desire of academics to have objective measures of their performance and that relative objectivity cannot be achieved without extensive formal peer review; the institution's own desire to understand the quality of their own portfolio and to use the data obtained for their own purposes, particularly reputational enhancement; the need to allocate a significant portion of higher education funding. In turn, the fact that funding is linked to the assessment, means that institutions take the assessment very seriously. These factors drive up the intensity of the process.
83. The dual-support funding system to HEIs is well embedded within the UK system and to undermine that at a time when other parts of the UK funding and support structure for research are uncertain would be unwise. We note the Nurse Report makes suggestions about reforming the block-grant funding system.
84. Critiques of the cost should recognise that all industries including academia require data to make the best strategic decisions and while academia is somewhat privileged with relative autonomy, the operational research

³⁵ Harnessing the Metric Tide: indicators, infrastructures and priorities for responsible research assessment in the UK Stephen Curry, Elizabeth Gadd and James Wilsdon. Report of The Metric Tide Revisited panel. December 2022. <https://s3-eu-west-1.amazonaws.com/pstorage-rori-0700695484/38515103/HarnessingtheMetricTideDecember2022final.pdf>

³⁶ www.sfdora.org

generated is required by those who fund HEIs. The intermediate information is also immense value to the institutions themselves.

85. The key question is how the cost can be minimised without obviating the value of the exercise. In many ways the burden is driven by the expectation of stakeholders because money and reputation are involved. It is the academic community and their managers who demand apparent objectivity for something that always has a subjective component. Any academic who has been either the recipient of or undertaken peer review recognises its limitations and variability.
86. While traditional metrics are only used to supplement peer review in a limited number of panels, they are undoubtedly extensively used by institutions in their selection of the portfolio and they do heavily influence the academic culture. Indeed, the focus on bibliometrics is generally accepted internationally to have distorted the academic endeavour. Thus the IAG believes that substantive attitudinal change to output assessment across the entire academic system is desirable. We make this point noting that some commentators have argued that the correlation between bibliometric measures and funding allocations is sufficient to obviate the need for a REF to allocate funds (missing the point that it has other purposes and ignoring the issues of disciplinary diversity). Others have pointed out that the outputs are largely peer reviewed anyhow. But not all peer review is the same, and peer review itself is increasingly compromised.
87. Core to a future REF are three core questions:
- a. Is the research of high quality?
 - b. Has the institution taken leadership in evolving a strategy to produce quality impactful and engaged research?
 - c. Is the institution adaptive to changing times?
88. The subsidiary questions are then: how will each of these questions be answered, what data, what narrative, what evidence will be required? In turn, this leads to the challenge of guidelines. Very tight guidelines to applicants will drive homogeneity, looser guidelines require those doing the assessing to be pluralistic and wise.
89. Research assessment should be focussed more on the processes that are likely to generate high quality research rather than the outputs themselves. Research by its very nature is uncertain and sometimes serendipitous and yet too much of the focus has been on the individuals and their outputs rather than the circumstances created by institutions in which researchers operate. We believe this requires rebalancing in the REF process itself.
90. Underlying this discussion is a mismatch which is fundamental. The dominant output focused approach looks at individuals and their retrospective performance, yet the funding derived from the assessment goes to the

institution for their strategic choices on how to move forward. This drives a behaviour around the individual which can be unhealthy. It is a primary reason for obviating the focus on the individual and reducing the focus on "REF-able outputs".

91. Globally the trend is away from standard bibliometrics with all their foibles and the promise of AI based and other new measures will not obviate the negative impacts on individuals and units. Therefore, the IAG is strongly of the view that the focus on future assessment exercises should dramatically reduce the weighting given to research outputs and enhance that given to the 'people, culture and environment' and the level of impactful engagement the institution has with a broad range of potential stakeholders from the community to the policy sector to the private sector. Assessing these elements must, to a significant extent, be subjective and require some form of panel approach, but submissions must be supported by evidence and some elements can clearly be metricised.
92. Any assessment must not seek to homogenise university research across similar disciplines or place academics in traditional disciplinary silos. Rather any assessment must seek to reward diversity and innovation and distinction between units similar in nature.
93. While revolution may be attractive to some, the IAG recognised the pragmatic realities of change but believes the next exercise should give weight to the research outputs at far less than 50%. Rather the focus should be on developing and using a range of narrative and metrics that reflects the requirement of society to see that research ranging from the most basic to a range of creative outputs serves the interest of society (broadly defined), that they are assured that HEIs are adapting and evolving to continue to maintain the UK at the forefront of academic research in all domains.
94. Core to progressing the REF is to simplify its purpose. One incentive scheme cannot possibly deal with at least six distinct and not necessarily aligned objectives (as detailed in the 2014 report). It might be that even six may not represent all the areas that might be incentivised – such as culture change or specific classes of activity (e.g., inter and transdisciplinarity). The IAG recommends simplifying the purposes of the REF to:
 - a. Informing the dual-support funding system and allocating funds to HEIs based on research quality to include the environment in which the research will be conducted (broadly defined)
 - b. Providing accountability to all stakeholders for public investment in research via HEIs
 - c. Provide data and insights that will support further enhancement of university research in the UK
95. With the first of these recommendations the devil is in the detail and the temptation could be to revert to the REF as it was in 2021. We recommend that

the focus shift further to measures of both ordinary and extraordinary impact and engagement and to the broader research-supporting environment within each institution. It should also shift as much as possible to the institution rather than individual while recognising that universities are ultimately a human community. If necessary or desirable, consideration could be given to using other fiscal tools to incentivise particularly desired and specific actions beyond these core purposes.

96. The Funding Bodies should explore the consequences of such recommendations to panel process simplification and to the removal of returning any measures that do not meet the core purposes of the REF as revised.
97. Further change is likely needed in the future (beyond 2028) recognising that some domains of academia are undergoing substantive evolution towards more team-based research, transdisciplinary research and more diverse actors. Further technological advances (e.g., in AI) and the milieu in which the UK strives to be a global leader in research will continue to change. Much reflection is underway across the global community about research assessment and peer review and the changes we suggest are likely only a part of a journey to further evolution in later years. In taking an evolutionary rather than a revolutionary approach we encourage the Funding Bodies to be thinking beyond the next assessment round to what might follow.
98. Finally, we must acknowledge that no incentive system is perfect. There will always be spill-over consequences and this periodic review is always needed. Sadly, it is institutional and human nature that any incentive system will be gamed by some. A level of misuse by institutions for example in staff management or in institutional reputational efforts is likely inevitable. But the REF is an important source of funding and is sustaining academic research in the UK in increasingly uncertain times. If used with goodwill and common sense, it can help ensure the UK remains a strong global power in academic research and its application across all disciplines to benefit the UK and its citizens.

Recommendations

1. That the basis of the REF be retained and continue to be linked to a dual-support funding system research with the UK's higher education facilities.
2. That the purposes of the REF are clarified and simplified in the context of the changing nature of academic enquiry. These would be:
 - a. To continue to provide a source of institutional funding to promote research linked to past and expected performance,
 - b. To give the UK and devolved UK Government's and UK taxpayers confidence that the system is accountable for use of taxpayer funds and will sustain and promote UK's global standing,
 - c. Provide system wide and institutional data for performance enhancement.
3. That, if desirable, consider other ways of incentivising other activities beyond these core purposes with dedicated funds thus clarifying the link between REF and HEI performance and both strengthening and weakening (depending on the specific objective) the link between REF and funding.
4. Value the processes and environments that increase the probability of generating high quality and impactful research more than the past outcomes.
5. Recognise that the nature of research is evolving. In that context any REF must now recognise team-based research and the contribution of a broad scope of individuals who have in the past not been assessed. It must recognise the growing and essential contribution of under-represented communities.
6. Define, support and develop better processes to incentivise and support inter- and transdisciplinary research. The issues here extend beyond research assessment to both funders and employers.
7. The balance between outputs (contribution to knowledge and understanding), impact (engagement and impact), and environment (people, culture and environment) needs to change. We recommend each should be weighted in the next round at 33%.
8. The definitions of outputs (contribution to knowledge and understanding), impact (engagement and impact), and environment (people, culture and environment) require revision with greater clarity of what is sought:
 - a. Redefine excellence/quality research in a more pluralistic manner
 - b. Reduce the system-wide focus on classic quantitative bibliometrics in assessing outputs
 - c. Break link to individuals
 - d. Seek ways to support distinctive profiles of units and true innovation in unit profiles
 - e. Impact must move from singular focus on exceptional impact to promote both exceptional and ordinary and sustained impact and stakeholder engagement

- f. The research environment should be assessed at the institutional level and include measures to recognise and reward positive research cultures and environments
 - g. Research outputs should be assessed at the unit level not the individual level
 - h. Transdisciplinary and interdisciplinary outputs and impact are likely to involve multiple units and potentially institutions and the evaluation process must recognise and incentivise this
9. Recognise that diversity and plurality in the range of academic research in the UK is critical as much as current outcomes. Therefore, panel membership and criteria should reflect the evolving nature of research and the changing purposes of the REF as revised
10. UK funding bodies and HEIs should collaborate to review what data are collected and the rationale for its use. UK funding bodies, funders and institutions should look to reduce the time burden placed on academics and the overall cost of the REF system.
11. HEIs should be encouraged to reduce the burden associated with their internal use of the REF and giving it excessive emphasis (for example talking about 'REF-able' and 'non-REF-able' research or researchers). The REF should not be used as a substitute for staff performance or other within-institution activities or seen as a ranking tool.
12. UK funding bodies, funders and institutions should monitor and evaluate the impact of the REF on research culture and the UK's research and innovation landscape as a whole.
13. Note that it is likely that further evolution of the REF will be needed after 2028. Any change in research assessment will have both intended and unintended consequences. There remain emergent challenges that this report has highlighted which will need further consideration.

Appendix A – IAG members

Name	Job Title & Affiliation	Representing
Chair: Sir Peter Gluckman	President, International Science Council, New Zealand Director Koi Tū, the Centre for Informed Futures, University of Auckland Chief Scientific Officer: Singapore Institute for Clinical Sciences.	Research User
Dr Lidia Borrell-Damián	Secretary General, Science Europe	Funder
Dr Chonnetia Jones	President and Executive Director, Addgene. Formerly Vice President, Research, Michael Smith Foundation for Health Research, Canada. Formerly Director, Insight & Analysis, Wellcome Trust, UK.	Funder
Professor Erika Kraemer-Mbula	Professor of Economics, DSI/NRF/Newton Fund Trilateral Chair in Transformative Innovation, the Fourth Industrial Revolution and Sustainable Development, University of Johannesburg, South Africa	Academic
Professor Jane Ohlmeyer	Erasmus Smith's Professor of Modern History (1762), Trinity College Dublin and Chair of the Irish Research Council (2015-21)	Academic
Dr David Phipps	Assistant VP Research Strategy & Impact, York University and Director of Research Impact Canada	Research Professional
Dr Kirsty Salmon	VP advanced bio & physical sciences for low carbon energy, bp	Research User
Ms Judi Zielke*	Chief Executive Officer, Australian Research Council	National Assessment Expert

* Ms Judi Zielke was preceded on the group by Professor Sue Thomas, who was CEO of the Australian Research Council until April 2022.

Appendix B – IAG meetings

Meeting date	Phase	Pre-meeting date	Content summary
17 May 2021	1	13 May 2021	IAG's ToR. Introduction to FRAP. Key topics to be explored/discussed: Definition of Research Excellence, Changing nature of academic enquiry and knowledge exchange, equity for career and discipline, granularity of assessment narrative versus metrics, community acceptability of changes.
29 June 2021	1	25 May 2021	Key points/emerging themes from roundtables and stakeholder engagement. Definition of research culture, incl. the overlap with research environment and link to excellence. Impact in REF and for FRAP
27 July 2021	1	22 July 2021	Further in-depth look at impact: issues relating to assessment of impact and how FRAP might minimise these. Research culture definition.
15 September 2021	1	18 August 2021	Assessment of research environment. IDR/TDR incl. definition and assessment
18 October 2021	1	6 October 2021	Professor Wilsdon's thoughts on FRAP-related issues in light of his work on Real-time REF review and Metric-tide in particular. In-depth look at Research Excellence.
2 December 2021	1	18 November 2021	Emerging themes from FRAP's evaluation and stakeholder engagement activities. Bringing together the Group's discussions so far and reflecting on where they got to and where next.
27 April 2022	2	10 March 2022	Summary of Chair's discussion with Lord Stern*. Update on FRAP's activities. Introduction into Board's considerations of options
27 July 2022	2	13 July 2022	The Group's reflections on where it got to, what it needs to achieve and how to approach Phase 2. Headline (quantitative) findings from FRAP consultation. Overview of EUA/SE agreement.
15 September 2022	N/A	N/A	Extraordinary meeting with the team working on Metric Tide Revisited.
26 October 2022	2	6 October 2022	Options analysis incl. drivers for change.
15 February 2023	2	7 February 2023	Programme Board's direction of travel for FRAP initial decisions. IAG's report: form and content.
3 April 2023	2	None	IAG's draft report and recommendations.

* Lord Stern met with the Chair of the IAG on 23 March 2022.