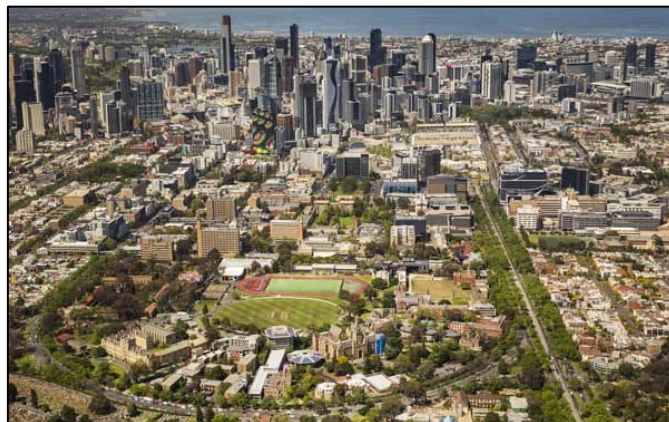




The University of Melbourne
Response to the
Australian Research Council
ERA EI Review Consultation Paper
October 2020



Introduction

The University of Melbourne welcomes the opportunity to contribute to this important review that seeks to enable the ARC to: simplify and streamline the programs; take advantage of recent developments in technology and big data; ensure the programs continue to reflect world's best practice; and respond to the ongoing needs of the university sector, government and the public for a robust evaluation of Australian university research quality, impact and engagement.

Our response has been informed by the views of many senior staff across the University who were active in the preparation of the University submission to ERA 2018 and EI 2018, many of whom also contributed to our submissions in previous ERA assessment exercises.

We are aware that the review process has generated a deal of interest across the sector in issues such as definitional matters, timeframes, the perceived effectiveness of the initiatives in achieving robust measurements of excellence, impact and engagement, and potential efficiencies – both for the ARC and for the universities.

We acknowledge the comprehensive nature of the Consultation Paper and the extensive list of consultation questions. We have not answered each one of these, aware that our views coincide with many others. We have sought to highlight some issues for consideration by the Review Advisory Committee,

We would be pleased to have the opportunity to elaborate on any matters.

Submitted on behalf of The University of Melbourne

*Liz Sonenberg
PVC Research Infrastructure and Systems
Chancellery Research and Enterprise
l.sonenberg@unimelb.edu.au
October 2020*

Section 3—Excellence in Research for Australia

Q3.1 To what extent is ERA meeting its objectives to:

- a. Continue to develop and maintain an evaluation framework that gives government, industry, business and the wider community assurance of the excellence of research conducted in Australian higher education institutions.
- b. Provide a national stocktake of discipline level areas of research strength and areas where there is opportunity for development in Australian higher education institutions.
- c. Identify excellence across the full spectrum of research performance.
- d. Identify emerging research areas and opportunities for further development.
- e. Allow for comparisons of research in Australia, nationally and internationally, for all discipline areas.

The ERA exercise is not fully meeting its objectives. In terms of the requested response options, suggest recording this as “*A moderate amount.*”

There are essentially two different systems in place – one that is citation based (largely STEM-M disciplines) and one that is peer review based (largely HASS disciplines) as has been noted by other commentators¹. Other analysis has shown that all citation-based disciplines are steadily improving their ratings, while ratings in peer reviewed disciplines tend to oscillate from year to year². Some have questioned whether the measured improvements are credible or are a consequence of improved submission preparation. Given an objective of ERA was to incentivise and recognise excellence, it may be timely to refresh the objective and reconsider the approach to assessment.

For example, introducing citation-based assessment in more disciplines, and reducing the overall volume of material to be handled and in particular, peer reviewed, perhaps by using a ‘4 best’ or similar (as in the UK REF).

With regard to identification of emerging areas, as the approach is retrospective it is not effective as a guide to changing trajectories. An alternate means of identifying emerging areas could be through an analysis of recent grant applications (awarded and not awarded) to selected granting bodies.

Q3.2 The ERA objectives are appropriate for meeting the future needs of its stakeholders.

Disagree. We acknowledge the national benefit of safeguarding public investment in research. However, the extent to which this can be achieved through more streamlined means than ERA has become urgent to consider. The value of ERA to guide institutional decision making is low – ratings are a blunt instrument, highly retrospective, typically not aligned with institutional organisational arrangements and lacking in local strategic context. While some commentators endeavour (against the advice of the ARC) to convert outcomes to league tables at a macro or discipline level, these are not meaningful. Additionally, the metric of “world standard” remains undefined and much of the benchmarking that is most helpful is around capability rather than FOR code.

Q3.3 What impacts has ERA had on:

¹ <https://odysseyhe.tumblr.com/post/629407253284421632/98-questions>

² https://franklarkins.files.wordpress.com/2019/05/a34.f-p-larkins_era-excellence.pdf

a. the Australian university research sector as a whole

While there has been an increase in rating values since the introduction of ERA, it is problematic to suggest that this is a measure of absolute increase in excellence, as noted by many commentators. It may be that the existence of a national assessment exercise has provided some assurance, at the level of national decision makers, about investment in university research. However, the lack of reach of such assessment to cover other publicly funded research organisations surely limits the interpretability of the exercises for national assurance purposes.

b. individual universities

This University's comprehensive publication dataset has been used for internal purposes at various points to understand our research profile including the nature and extent of collaborations within and external to the University. However, the University can conduct the same analysis independent of ERA, and the availability of new databases and digital tools makes this increasingly straightforward an exercise.

c. researchers

None. Feedback indicates no effect on individual researchers, who are consistently advised to take discipline specific input on publication strategy.

Q3.4 How do you use ERA outcomes?

We do not use ERA outcomes in institutional decision making. We have demonstrated that we can acquire data on comparative performance from other sources.

Q3.5 ERA outcomes are beneficial to you/your organisation.

Disagree. See above

Q3.7 The current methodology meets the objectives of ERA.

Disagree. See above

Q3.15 The peer review methodology for evaluating the quality of research is appropriate.

Neither agree or disagree. It is accepted that there are disciplines where citation analysis remains inappropriate, but there are arguments for expanding the methodology to incorporate more HASS disciplines that are still heavily based on journals; for example, history, sociology, political science, etc could all use citations. Others have elaborated on lessons to be learned about the usefulness of peer review compared to the effort involved³.

Q3.19 The volume and activity indicators are still relevant to ERA.

Agree.

Q3.20 The publishing profile indicator is still relevant to ERA.

Agree.

Q3.21 The research income indicators are still relevant to ERA.

³ <https://harzing.com/publications/white-papers/running-the-ref-on-a-rainy-sunday-afternoon-do-metrics-match-peer-review>

Agree.

Q3.22 The applied measures are still relevant to ERA:

a. Patents.

Strongly disagree. It is heavily biased to STEM disciplines and is not correlated with the quality of the research produced.

b. Research commercialisation income.

Strongly disagree. It is heavily biased to STEM disciplines is not correlated with the quality of the research produced.

c. Registered designs.

Strongly disagree. It is heavily biased to design disciplines and is not correlated with the quality of the research produced.

d. Plant breeder's rights.

Strongly disagree. It is heavily biased to botanic disciplines and is not correlated with the quality of the research produced.

e. NHMRC endorsed guidelines.

Strongly disagree. It is heavily biased to Medical disciplines and has is not correlated with the quality of the research produced.

ERA rating scale

Q3.23 The five-band ERA rating scale is suitable for assessing research excellence.

Neither Agree nor Disagree. While we appreciate the Review Committee considering this fundamental question, we believe this to be a question to which there are many facets to a possible answer. In part they link to the difficulty of a linear scale to measure multidimensional qualities. In part it points to question 3.37 of signalling volume in the assessment outcomes, as an intensity indicator.

Q3.24 Noting that 90% of units of evaluation assessed in ERA 2018 are now at or above world standard, does the rating scale need to be modified to identify excellence?

It remains a fundamental question for the Review to determine whether changing the rating scale would bring useful discriminatory power, or perhaps whether the initial objectives of the ERA exercise have now been met and it should be changed to reflect other priorities.

Q3.25 The ERA low-volume threshold is appropriate.

Strongly disagree. For the University of Melbourne, 50 outputs represents less than 0.1% of the research conducted so is an inappropriate threshold when the outcomes are compared, by discipline, across institutions of substantially different scale.

Q3.26 Are there ways in which the low-volume threshold could be modified to improve the evaluation process? *Please describe.*

A sliding scale referenced to the size of the institution.

Q3.27 What is the more appropriate method for universities to claim research outputs—staff census date or by-line?

They allow the capture of different ambitions of an assessment exercise – so the choice is dependent on the objective.

Q3.28 What are the limitations of a census date approach?

We are all aware of instances where staff have arrived at an institution just before the census so their research was ERA eligible and hence reported, but realistically there was no association with the institution prior to the exercise, and in some cases neither was there sustained subsequent contribution at that institution. Additionally, we note that with a census date approach there can be a major load associated with coding previous outputs of staff as they join a university, which reduces the opportunity for streamlining.

Q3.29 Would a by-line approach address these limitations?

Yes.

Q3.30 What are the limitations of a by-line approach? *Please describe.*

They allow to capture different ambitions of an assessment exercise – so the choice is dependent on the objective.

Q3.31 ERA adequately captures and evaluates interdisciplinary research.

Strongly disagree. By its definition, ERA is a discipline focused exercise. Further, in the citation-based disciplines, the methodology of proportioning citations actively discourages multicoding of FOR codes to research outputs.

Q3.32 My institution would meet ERA low-volume threshold in Indigenous studies at:

a. Two-digit?

We note there is only one two-digit code.

Importantly, we strongly believe that the proper question at this point is whether the University 'could' meet a threshold (i.e. not 'would') – and yes, it could. As prior work has already been coded under an alternate regime of disciplinary considerations, detailed internal discussions would need to be conducted and an approach agreed upon before a decision is made on the use of the new codes.

b. Four-digit?

We estimate we could meet the four-digit threshold in multiple codes, but please also refer to the response to 3.32 a.

Q3.33 In ERA, the best approach for evaluating Indigenous Studies is (choose one):

a. Using established ERA methodology i.e. the low-volume threshold would apply to the Indigenous Studies discipline and all its specific disciplines

b. For Aboriginal and Torres Strait Islander studies by combining low-volume disciplines into single units of evaluation

c. For Aboriginal and Torres Strait Islander studies by combining low-volume disciplines into two units of evaluation (one unit comprising Humanities, Arts, and

Social Sciences disciplines and one unit comprising Science, Technology, Engineering and Mathematics disciplines)

d. Other. Please describe.

Indigenous studies as described in the new ANZSRC incorporates a wide range of disciplines. There is some concern whether there will be sufficient expertise in a 45 panel to evaluate the breadth of disciplines represented. Further, for ERA 2023, as much of the work to be represented in the 45 code will have previously been coded under the previous disciplinary regime, there will be a major manual workload in reconsidering allocation under the new codes.

For the next ERA, assuming it takes place in 2023, the University of Melbourne suggests the 45 codes should be deployed *in addition to* the other codes and hence used in a pilot way and to identify and profile (not assess) the breadth and depth of Indigenous research being conducted across the sector.

Q3.34 What would be the advantages and/or disadvantages of your preferred approach for evaluating Indigenous studies in ERA? *Please describe.*

Making the 45 codes additive to the other codes in 2023 would allow the ARC some piloting of the approach, with less 'at stake' and also timely analysis of the indigenous research being conducted. A staged approach, towards ERA 2028 would allow the coding (of publications and income) to occur prospectively, not retrospectively, and be a more authentic representation. The complexity of getting the data 'right' (as illustrated by some responses to recent analyses⁴), suggests a staged approach across the next two ERA exercises would have considerable advantages.

Q3.35 ERA should move to an annual collection of data from universities.

Neither agree or disagree.

Q3.37 In future ERA rounds, should the volume of outputs submitted for each unit of evaluation be included in the National Report?

Yes. Understanding the volume of research conducted in disciplines gives a better picture of where research is being conducted across the sector as well as where the quality lies. It also puts the ratings into context.

Q3.38 In future ERA rounds, research outputs should be published with their assignment to specific disciplines following completion of the round.

Strongly disagree. This will potentially result in conflict between institutions and academics on how their research was coded. There is no benefit to making this data available publicly.

Q3.39 What other data do you think the ARC should publish following an ERA round? *Please describe.*

None.

⁴ https://clarivate.com/webofsciencegroup/campaigns/anz_indigenous_research_report_2020/

Section 4—Engagement and Impact Assessment

EI Overview

Q4.1 Considering that EI is a new assessment, to what extent is it meeting its objectives:

- a. encourage greater collaboration between universities and research end-users, such as industry, by assessing engagement and impact?
- b. provide clarity to the Government and the Australian public about how their investments in university research translate into tangible benefits beyond academia?
- c. identify institutional processes and infrastructure that enable research engagement?
- d. promote greater support for the translation of research impact within institutions for the benefit of Australia beyond academia?
- e. identify the ways in which institutions currently translate research into impact?

The Review Committee will be very familiar with the many commentaries and critiques of ways to incentive, measure and evaluate the behaviours that EI seeks to capture⁵. While acknowledging the importance of ensuring strong attention is paid to mutually beneficial interaction with external parties and research translation, and impact beyond the academic communities, we are not persuaded that the current construction of the EI exercise achieves this. A fundamental question is whether the indirect impact of changing culture by (retrospectively) assessing outcomes can achieve as much as other approaches that might focus on universities having structures in place to support engagement and impact, and hence more directly incentivise modified behaviours.

Q4.3 What impact has EI had on:

- a. The Australian university sector as a whole? *Please describe.*
- b. Individual universities. *Please describe.*
- c. Researchers. *Please describe.*
- d. Other sectors outside of academia? *Please describe.*

Minimal.

Q4.4 How do you, or your organisation, use EI outcomes? *Please describe.*

The University has made major investments in research engagement and impact on all their dimensions, but these have been the outcome of deliberate strategic choices pre-dating the EI exercise. The EI impact studies have been deployed, as have many many other stories and studies that we generated independently of the EI assessment, and much more economically, as part of our overall research profiling. The EI rating outcomes have not been used.

⁵ (e.g.) *The Impact Agenda: Controversies, Consequences and Challenges*. Katherine E. Smith, Justyna Bandola-Gill, Nasar Meer, Ellen Stewart and Richard Watermeyer. Policy Press. 2020. (e.g.) *The Evaluators Eye*, Gemma Derrick, Palgrave 2018. (e.g.) Terama, E., Smallman, M., Lock, S. J., Johnson, C., & Austwick, M. Z. (2016). Beyond Academia - Interrogating Research Impact in the Research Excellence Framework. *PloS one*, 11(12), e0168533. <https://doi.org/10.1371/journal.pone.0168533>

Q4.5 The EI outcomes are valuable to you or your organisation.

Strongly disagree. See 4.4.

Q4.7 - Q4.10 on definitions

See responses to 4.1, 4.3, 4.4.

Q4.11 Are the two-digit Field of Research codes the most appropriate method to define units of assessment for Engagement and Impact?

Yes, given other limitations and sector familiarity with them, and now the changing use of SEO codes for ABS reporting.

Q4.13 Should there be more or fewer units of assessment per university?

The number should be no greater than in EI 2018 – and as below, without the interdisciplinary impact study.

Q4.16 Overall, the engagement indicator suite for the assessment of research engagement is suitable.

Strongly disagree. Applied measures are very STEM focused, and are exceedingly narrow indicators of engagement.

Q4.17 The cash support from research end-users indicator using HERDC data is appropriate for the assessment of research engagement?

Strongly disagree. Cash support from End Users is an input, and while quantitative is not a proportionable measure of engagement.

Q4.18 The research commercialisation income is appropriate for the assessment of research engagement.

Strongly disagree. Research commercialisation outcome, while quantitative, is not a proportionable measure of engagement as there are so many confounding factors.

Q4.21 Should any of the current Engagement metrics be redesigned?

The EI endeavour to use available metrics as a proxy for Engagement, rather than inventing an extensive suite of purpose-built measures that would be expensive and possibly infeasible to collect systematically, was appreciated. However, in practice and on reflection, we do not believe a meaningful and useful set of indicators was identified – and believe that redesign by looking at alternative metrics, rather than considering a fundamentally different assessment exercise, is bound to be unsuccessful for similar reasons.

Q4.22 The co-supervision of HDR students should be made an engagement indicator in future rounds of EI.

Strongly disagree. While many candidates have wonderfully positive experiences through external involvement, co-supervision is only one aspect of how this can be conducted to the candidate's benefit. Co-supervision also ignores the situation of the graduate research candidate who may not wish to have industry co-supervision, so establishing a baseline of "appropriate percentage" of co-supervised candidates in a discipline would become a curious challenge. Overall this would be an overly narrow metric, relatively easy to increase numerically without achieving authentic positive consequences for the candidate. Creating incentives for superficial 'engagement' would yield a potentially perverse outcome.

Q4.23 In your opinion, are any of the ERA applied measures appropriate indicators of research engagement in EI?

a. Patents.

No. These metrics are heavily weighted to STEM disciplines and do not necessarily reflect engagement. A university could patent a technique with no external engagement.

b. Research commercialisation income.

No. As above.

c. Registered designs.

No. This metric is only applicable to one discipline – designs can be registered independent of having been the result of engagement.

d. Plant breeder's rights.

No.

e. NHMRC endorsed guidelines.

No.

Q4.24 The narrative approach is suitable for describing and assessing research engagement with end-users.

It is clear from our earlier responses that the University does not support the use of quantitative indicators for Engagement. If there continues to be a separate Engagement assessment, the ARC may wish to consider whether to prioritise the collection of information over the flow and polish of the narrative itself. Narrative writing demands the collection of information. It also requires the synthesis of that information into a highly organised, linear sequence, which is pleasing to read - the process is time-consuming and depends on particular skill set. Narrative writing is typically used for persuasive purposes, or to share highly complex and nuanced information. How easily the reader can discern the information from the narrative depends more on the quality of the writing than on the information that the writing contains. Replacing the narrative with a Question and Answer format might reduce the workload for both those involved in submission and assessment. These questions could encapsulate the prompts that have been used to guide completion of previous engagement narratives, as well as extract fresh information ARC wishes to obtain.

Q4.28 The narrative approach is suitable for describing and assessing impact.

Agree.

Q4.29 One impact study per broad discipline is sufficient for capturing the research impact within that discipline.

Neither agree or disagree. The primary value of the exercise is the very impressive aggregate collection of impact studies which contribute to national understanding of the value, outside academia of national research investment. A uniformly collected body of work, aggregated at the national level, has potential power. But it should not be seen a tool for institutional PR, or discipline comparison, or rankings or league tables – such endeavours are not meaningful, whether they be derived from a single impact study per unit of evaluation per institution, or a greater number. We suggest that there could be opportunity for pooled case studies between institutions, as this reflects the realities of how research actually

occurs, and could enable a more informed public discourse on the nature and impact of research for national benefit.

Q4.32 In your opinion, are there quantitative indicators that could be used to measure the impact of research outside of academia?

No.

Q4.33 The narrative approach is suitable for describing and assessing approach to impact.

Neither agree or disagree. The University suggests integrating 'Approach' with the impact narrative, but not treating it as a separable item meaningful at the two-digit level. There are no metrics that can explain the concept of an approach to impact as it is essentially a qualitative concept. Consistent with our view that the value of the Impact exercise is the aggregate body of work at a national level, one could consider one Approach statement for the University (not per case study). That public description would provide a form of incentive, if one is perceived to be needed, for universities to profile their extensive endeavours in supporting and enabling the achievement of impact, without the burden on the ARC of seeking to compare this, via ratings, across institutions.

Q4.38 to Q4.43

Neither agree or disagree. We are interested to hear the views of those involved in the assessment and the ARC's own experience in how the ratings are interpreted in the national discourse, outside academia, about the exercise, and contribute to a subsequent consultation on these matters, once the broader ambitions of the present Review conclude.

Q4.44 Should EI continue to include an interdisciplinary impact study in addition to the two-digit Fields of Research impact studies?

No. Essentially all of the University of Melbourne impact studies were interdisciplinary and could have been assessed in any number of two digit FoR codes. It is redundant to have an interdisciplinary case study.

Q4.45 Should the EI low-volume threshold be applied to the unit of assessment for Aboriginal and Torres Strait Islander research in EI 2024 with the option to opt in if threshold is not met?

Please refer to our responses to items 3.33 and 3.34

Q4.46 Should the unit of assessment for Aboriginal and Torres Strait Islander research include engagement in EI 2024?

No. Data issues need first to be addressed.

Section 5—Overarching Issues Common to both ERA and EI

Q5.1 How often should ERA occur?

Every five years. It would reduce the workload on an already overburdened higher education sector and not reduce the reliability of the assessments.

Q5.2 What impact would a longer assessment cycle (i.e. greater than three years) have on the value of ERA results, particularly in the intervening years? *Please explain your answer.*

As there is relatively little volatility in results, a longer cycle does not appear to be problematic.

Q5.3 How often should the EI assessment occur?

Every five years. It would reduce the workload on an already overburdened higher education sector and not reduce the value of the assessments.

Q5.4 What impact would a longer assessment cycle (i.e. greater than three years) have on the value of EI results, particularly in the intervening years? *Please explain your answer.*

As a tool for raising the awareness of the important contributions of research to the community and the nation, the presence of the exercise is arguably more important than the frequency.

Q5.5 ERA and EI should be combined into the one assessment.

Strongly Disagree. The workload involved is already substantial and combining it into one time period will overload an already overburdened and potentially smaller workforce post COVID.

Q5.6 Are there other ways to streamline the processes to reduce the cost to universities of participating in ERA and EI?

Yes. Reuse sources wherever possible.

Q5.7 In your view, what data sources could ERA utilise?

SciVal, ARC/NHMRC Grant Databases, Metadata from Scopus/Web of Science

Data from tools like SciVal could be used to assess the citation-based disciplines. Our internal analysis has shown a high correlation between ERA Ratings and Field Weighted Citation Impact.

Grant results from both the ARC and NHMRC could be used in the analysis given a) FoR codes are provided when applying for funding and b) funding is reported on a per year basis.

Metadata should be sourced from the citation provider – universities should not be required to provide data that the ARC already has access to.

Q5.8 In your view, what are the most time-consuming elements of an ERA submission?

There is a huge amount of metadata that is reported that the ARC already has and offers no value in terms of assessing the output. This includes listing all authors on a paper (for a large ATLAS papers this runs into 3-4,000 authors at time), the place of publication (meaningless data given publishing houses are often based across multiple countries - this should be removed as a required metadata field), reporting research income at the grant level when the ARC already has most of this data and it is reported at a consolidated level to the

Department of Education. These data elements create a huge amount of work for little benefit.

There are also a number of business rules which are confusing and do not add any benefit to the value of the assessment.

BR059 (65/35 rule) is arbitrary and bears no resemblance to how researchers code their output. It is also confusing to explain and generates an error rather than a warning. It doesn't help in any way in terms of appropriately coding research and to some extent undermined the trust that academic staff have in the ERA process.

BR113 is similar and again bears no relationship to how academics code their research. This also adds a huge additional burden onto universities for no additional value in terms of the assessment.

BR104 requiring institutions to list all creators/authors on an output is incredibly time consuming (particularly for large physics and public health papers) and offers no value in terms of the assessment.

BR106 should be aligned to how the income is reported in ERA (more so now that ACGs are reported with codes to the HERDC – the same codes should be used in ERA). Also separating the codes by year when only one code is required to describe the project is not required.

BR047 – there are instances where the words “Not available” are legitimate in an ERA context including being part of an explanatory statement. NA is also the chemical symbol for sodium and we did have instances where we received an error message where capitals had been used in output titles.

Q5.9 In your view what are the most time-consuming elements of an EI submission?

a. Are there efficiencies that could be introduced?

Yes. Essentially the same research income data that is reported for ERA is reported for EI. The data should all align. The same categories that are reported for the HERDC should be used for both ERA and EI.

Q5.10 ORCID IDs should be mandatory for ERA.

Disagree. The practice-based disciplines have very poor uptake of ORCID. It could work for more traditional disciplines but will require a substantive startup process, to ensure coverage – and so could only be managed with appropriate lead time. Another issue would then be how would FoR codes be assigned to outputs if they are directly drawn down from ORCID.

Q5.11 The automatic harvesting of output data using ORCID IDs could streamline a university's submission process.

Disagree. It is unclear how FoR codes would be assigned in this instance. It would also require matching up what is in ORCID with what is in our enterprise publication system. It would require 100% compliance with all researchers in order to be feasible.