Submission

Australasian Council of Deans of Arts, Social Sciences and Humanities

Australian Research Council review of the Excellence in Research Australia (ERA) 2023 Benchmarking and Rating Scale

April 22, 2022

The Australasian Council of Deans of Arts, Social Sciences and Humanities appreciates the opportunity to contribute to this consultation process around the Australian Research Council review of the Excellence in Research Australia (ERA) 2023 Benchmarking and Rating Scale.

A previously published review into ERA and the Engagement and Impact Assessment (EI) recommended changes to this process and the Acting Minister's recently published Letter of Expectations to the ARC CEO highlights the need to expedite the process to provide clarity to the sector.

Below you will find our full response to each of the questions outlined in sections 2, 3 and 4. In addition, DASSH introduces a number of questions generated through consultation with our membership.

We acknowledge the ARC is still consulting with Indigenous academics regarding guidance about Indigenous research, and that the Universities Australia Indigenous PVC/DVC group is also providing advice about these matters. We share sector-wide concern about the utility of the peer review criteria which we have noted, specifically about who will conduct assessments and how Indigenous research will be assessed, given the relative paucity of Indigenous expertise due to systemic inequities in the higher education sector.

DASSH recommends that Option A be adopted over Option B as it produces a better result in line with the review's recommendations. We see it as a more accessible qualitative descriptor. We believe that it meets the objectives of the review process because it 'cleans up' the lower end of the ratings scale, while also providing more distinctiveness at the top end of the scale. In addition, we note that Option A is somewhat similar to the existing ERA methodology and therefore more likely to result in outcomes aligned with the objectives for ERA nationally than Option B.

However, DASSH members want to see some distinctive changes made to the peer review process for the next round in 2023 whether Option A, B *or neither* are adopted.

First, we believe the *trial of a citation metric alongside the peer review process* would be beneficial. It could be captured in the dashboard for each UoE and visible to REC panel members only for a rounded assessment of the UoE as a pilot of the validity of the citation metric for disciplines in our panel(s).

However, peer review and citation-based assessments need to be improved upon in order to depoliticise the assessment process. Right now, peer review allows methodological disputes within disciplines to influence evaluations. For the humanities, arts and social sciences (HASS) to thrive under either of these options, it is recommended that peer review and citation assessment be combined, and that the ARC seeks to extend international input for assessments. Our members believe the assessment process will be more robust with greater input from international colleagues. This is particularly important given the HASS disciplines must be able to compete with citation-based disciplines. While citations remain an important part of the assessment of research overall in the sector, it must be noted that HASS researchers are also referenced through quotations in books, for example, which do not always appear in citation engines which tend to privilege digital/online publications.

Additionally, our members are clear that those conducting reviews should be senior staff with significant experience. DASSH proposes that all peer reviewers should be senior scholars such as Level D and E academics, given the proposed changes and the potential complexity of assessment.

We do not believe Option B is workable for peer review disciplines. Our members have concerns that Option B encourages gaming and tactical coding that will advantage some disciplines at the cost of others through the use of the proposed citation process.
In addition, the 'high performer' benchmark on its own is hard to determine objectively in peer review disciplines.

Regarding peer review guidance, our members have asked whether there will be separate guidance for Non-Traditional Research Outputs (NTROs). Several areas of HASS disciplines have largely NTRO outputs and much clearer guidance is needed in relation to the evaluation of these bodies of work. The REC panels dealt with these in the previous assessment but proposed changes to the assessment process need to consider offering specific guidance around this issue in light of intensified interest in cross-cutting research with applied outcomes in the HASS and creative fields.

DASSH believes there is a transition cost to moving in either direction. It will take time to disseminate the implications and develop understanding of the shift, however, there could be a significant negative consequence associated with either of the new scales.

As such, we would also support the suggestion of rejecting both Option A and B. In a process where there are so many FoR code changes, it may be wise to leave the rating scale as unchanged. Regardless of which option is adopted, the fact that ERA2023 ratings will not be directly comparable with previous ratings will have a negative impact on universities attempting to improve their performance based on previous ratings.

DASSH notes the Acting Minister's Letter of Expectation outlined a desire for industry engagement and impact:

"My expectation is that this should include clear measures to identify industry engagement and the translation of research to impact. I note that the Review identified many measures related to industry and end-user engagement as having a declining relevance to the assessment of quality in ERA."

DASSH Members are concerned that directions around 'impact' and 'industry engagement' will continue to disadvantage our disciplines because the nature of our impact and relations with the end users of our research are both different from that of STEM disciplines and also changing in a dynamic context for higher education. While the intervention can be seen as a 'call to action' for the sector – particularly in the creative arts – our role is to express some concern that the HASS disciplines are being sidelined through a direct process of industry alignment where there are no obvious connections for our disciplines.

We remain attentive to the prospect of end users becoming grant assessors at the ARC and we see any attempt to align the benchmarking system with the Federal Government's intention to commercialise research as a deeply concerning and flawed exercise. This is because such grant assessors are non-experts in research, may have conflicts of interest, and may also evaluate such matters based on their narrow interests and not on the broader interests of research and knowledge creation.

President Professor Catharine Coleborne
Australasian Council of Deans of Arts, Social Sciences and Humanities
Head of School / Dean of Arts
School of Humanities, Creative Industries and Social Sciences
College of Human and Social Futures
The University of Newcastle

Section 2: Options for a more granular rating scale

Q2.1 Which rating option (A or B) is preferred?

- We believe that Option A is preferable to Option B as it produces a better result in line with the review's recommendations. We see it as a more accessible qualitative descriptor. We acknowledge that it meets the objectives of the review process because it 'cleans up' the lower end of the ratings scale, while also providing more distinctiveness at the top. We note that A is more similar to the existing ERA methodology and more likely to result in outcomes aligned with the national interest than Option B.
- We note that A is more similar to the existing ERA methodology and more likely to result in outcomes aligned with the national interest than Option B.
- We do not believe Option B is workable for peer review disciplines, because there is not enough consensus in peer review disciplines about what 'world leading research' excellence looks like. In addition, the high performer benchmark on its own is hard to determine objectively in peer review disciplines.
- Option B also encourages gaming and tactical coding that will advantage some disciplines at the cost of others.

Q2.2 Are there particular features of either option that should be adopted or modified?

- Further clarity on the distinctiveness of each category would be welcome in the next level of detail following adoption of the scale. This would need to address the border between each category.
- Regarding Option A which collapses the old ERA ratings of 1 and 2 into one, 'Not at world standard' should be adopted – it is more important to know whether the research is above or below world standard than the level to which it falls below world standard.
- Regarding Option A, expanding the old ERA rating of 5 into two, 'World leading' and 'Well above world standard' should be adopted as it will better identify universities performing at the highest level internationally.
- Expanding the old ERA rating of 5 into three levels (as per Option B) is an unnecessary complication, particularly for peer review disciplines when determining the difference between UoAs is more subjective than in citation disciplines; it would be extremely difficult to frame questions that will prompt reviewers to consistently distinguish between such fine-grained levels of excellence.
- 'World standard' language should be used across ERA for consistency and clarity.

Q2.3 How will the change in ratings shift university research efforts?

- Once established, we expect there to be further efforts by research leaders to emphasise quality and excellence over quantum of output. For example, a furthering and deepening of existing settings moving away from HERDC quantum and towards specialisation and top 3 per cent type performance.

- Ratings will reinforce universities' emphasis on high quality outputs.
- We believe changes in rating scales will not have a material effect on researcher efforts.
- However, there is a risk that without tangible consequences relating to funding and the like, ERA scores risk having little resonance beyond the sector and already are not meaningfully understood internationally.

Q2.4 To what extent would the proposed options be more challenging for universities than the existing ERA rating scale?

- It will take time to disseminate the implications and develop understanding of the shift however there could be a significant negative consequence associated with either of the new scales.
- It has also been noted that the proposed changes to the peer review process might produce sector inequities in the now more diverse landscape of institutions and research UoAs.
- Option B would be a particular challenge for peer-review disciplines in distinguishing between the three levels, A, AA and AAA.
- Regardless of which option is adopted, the fact that ERA2023 ratings will not be directly comparable with previous ratings will have a negative impact on universities attempting to improve their performance based on previous ratings.
- The changes proposed to the rating scale aim to provide more detail at the higher end of the scale which effectively lower the rating of many UoEs for ERA 2023.
- We have concerns about the challenges that might result from these proposed options as they are likely to result in decreased ERA ratings (or the perception of decreased ratings) in all but the most outstanding handfuls of UoEs, and will be problematic in UoEs which have less volume in smaller universities.

Q2.5 What changes, if any, are required to the characteristics that accompany each rating level?

- Further clarity on the distinctiveness of each category is required along with the border between each category
- More detailed delineation between each level would be welcome, whichever option is adopted.
- How is the 'average standard of universities worldwide' being determined, particularly with reference to peer review disciplines? There is a need to set international benchmarks clearly based on evidence-based metrics.
- More work needs to be done on peer review indicators and guidance for peer reviewers.

Q2.6 Would it be feasible for expert reviewers to draw meaningful distinctions between each rating points using the characteristics provided?

- Yes, assuming further clarity on the distinctiveness of each category is forthcoming in the next level of detail following adoption of the scale.
- We think the ARC should insist on senior peer reviewers (Level D academics and above) given the proposed changes and complexity of assessment

dassh 20th anniversary

Collaborative
Influential
Connected
Knowledgable

dassh.edu.au

- We also suggest that either rating scale will be very difficult to operationalise in peer review. FoRs which lack quantitative metrics to allow objective measurement of the different ratings. Option B is too detailed at the top level without clarity about why this is needed when Option A also provides increased granularity.

Q2.7 What kind of additional training or guidance may be required in ERA 2023 to support the revised rating scale?

- Given that in previous ERA assessment rounds many Australian institutions have performed above or well above the world standard in many different FoRs, there will need to be a broad cultural change regarding the lack of comparability between ERA rounds not only in the university sector but more broadly amongst media, government, and industry
- On the issue of peer reviewing our members are clear that those conducting reviews should be senior staff with significant experience. This could be managed by stipulating that Associate Professors should be the most junior level reviewers.

Section 3: How can the citation metrics support the options for a revised rating scale?

Q3.1 How appropriate is the HPI as a method of supporting the rating scale options?

- Using only the top 10 per cent of organisations across the world is too limited, particularly given the inclusion of various types of UoE (such as universities as well as major scientific and other institutions); we would like to see this increased to 20% and limited to universities only to provide a more robust benchmark for comparability.

Q3.2 How appropriate are the dynamic RCI classes as a method of discipline-specific benchmarking?

- The dynamic RCI could be a useful reflection of differing citation practices between FoRs/disciplines.

Q3.3 How would the proposed citation methodologies impact research planning?

- As discussed above, DASSH members believe the citation and peer review methodologies need to be combined in a trial. As such the impact the proposed methodologies would have on research planning cannot be declared.

Q3.4 Do the new citation metrics support the drive for increased performance (especially in already high-performing disciplines)?

Q3.5 Is any additional criteria or information required in the citation disciplines to support the ratings at the highest end of each rating option?

- Our members maintain their concern at the need to identify 'citation disciplines' as distinct from peer review disciplines and reiterate that the citation assessment will put HASS disciplines at a disadvantage. This is because our disciplines typically comprise books and book chapters, which tend to appear less in citation metrics/indices, as well as peer reviewed articles and digital outputs that do attract citation metrics. The diversity of methods, approaches and forms of HASS scholarship means that peer review has to cover a much more diverse terrain. It needs to be able to provide the flexibility and sensitivity to express that and recognise diverse forms of excellence in ways that quantitative analysis is unable to do.

Q3.6 Please provide any additional comments on the proposed citation methodology.

- We believe the trial of a citation metric alongside the peer review process would be beneficial. It would be captured in the dashboard for each UoE and visible to EC panel members only for a rounded assessment. This would level the assessment process for both citation and peer reviewed disciplines. In this way the outcomes for universities favouring either STEM or HASS disciplines would also be levelled.

Section 4: How can the peer review indicator support the options for a revised rating scale?

Q4.1 To what extent are the proposed changes to peer review guidance likely to result in reports that are useful, informative and relevant for assessment panels? Please comment on any improvements that could be made, particularly with reference to disciplinary inclusivity and relevance to the options for the revised rating scale.

- Clarity on what constitutes quality of output in terms of world standards and benchmarks will be essential in helping engender clear and informative reports for panels
- Peer reviewers will need special and specific guidance, particularly in establishing benchmarks and distinguishing between upper levels of excellence.

Q4.2 How feasible would it be for peer reviewers to address the proposed peer review guidance? Please comment on any improvements that could be made, particularly with reference to clarity and workload for reviewers.

- The workload would be broadly similar to 2018.
- Guidance appears to be a work in progress so it is premature to judge how feasible it will be for peer reviewers.
- The dot points used to assess the contribution of each output are practical and should contribute to good reports from peer reviewers.

Q4.3 How appropriate is the proposed guidance for Indigenous studies? Please comment on any improvements that could be made.

- We recommend the use of some case study examples showing how these criteria could be expressed in different cases would be useful given this is the first time the codes will be used.
- Indigenous studies guidance seems appropriate – presumably it will continue to be refined in consultation with Indigenous assessors and will be supported by practical examples.
- We understand that the ARC is still consulting with Indigenous academics on this aspect, and that the UA Indigenous PVC/DVC group is also providing advice.
- We share concern about the utility of the peer review criteria which others in the sector have noted.
- We share the concern about how and who will be assessing, given the scarcity of Indigenous expertise.

Q4.4 How would the proposed changes to peer review guidance impact universities and/or researchers?

- They will further accelerate specialisation and excellence efforts - and further exacerbate inequalities between Go8/leading research groups and others.
- The methodology remains the same.
- The questions prompting peer reviewers to flag the presence of world-leading research may cause some confusion as to their precise meaning.
- How do reviewers judge what the disciplinary characteristics are of world-leading research in each field in order to establish benchmarks against which to assess outputs?
- Will there be separate guidance for Non-Traditional Research Outputs?

Q4.5 Is any additional criteria or information required in the peer review disciplines to support the ratings at the highest end of each rating option?

- Yes - global benchmark differentiators.
- As there were virtually no '1's given by peer review panels, it makes sense to merge 1 and 2. 4 is a big category and a bit amorphous which needs clearer criteria
- Examples for peer reviewers must be included, regardless of which Option is implemented. Guidelines should include a section for Non-Traditional Research Outputs.
- Examples of outputs to demonstrate ratings at all levels.

Q4.6 Are there any other changes to peer review that the ARC should consider?

- To reiterate, we are concerned there is no way to ensure that "the revised benchmarks will be interpreted consistently across the peer review and citations analysis methodologies". How would this be measured and implemented? Once again, we suggest the trial of a citation metric alongside the peer review process would be beneficial.
- We see a risk here that the nuances of much HASS scholarship will be lost as it is so deceptively simple to analyse citations. The diversity of methods, approaches and forms of HASS scholarship means that peer review has to cover a much more diverse terrain. It needs to be able to provide the flexibility and sensitivity to express that and recognise diverse forms of excellence in ways that quantitative analysis is unable to do. Indigenous research is one example of an area that could look 'weak' in a citations approach, but there are many other examples in our fields.