

DASSH¹ Response to ARC ERA EI Consultation

Table of Contents

<i>Excellence in Research for Australia</i>	2
ERA policy	2
Value of ERA.....	2
ERA methodology	7
ERA methodology at a glance	7
Citation analysis methodology.....	7
Peer review methodology.....	8
Contextual indicators	10
ERA rating scale.....	11
ERA low-volume threshold	11
ERA staff census date.....	12
ERA interdisciplinary research and new topics.....	12
ERA and Indigenous research	13
ERA process	13
Collection of ERA data.....	13
Publication of ERA data.....	14
<i>Engagement and Impact Assessment</i>	14
EI Overview	14
EI definitions	16
EI methodology	16
Unit of assessment.....	16
Selectiveness of EI.....	17
EI low-volume threshold	17
Engagement indicators	17
Engagement narrative.....	19
Impact narrative.....	20
Approach to impact Narrative	21
EI rating scales.....	21
EI interdisciplinary research.....	22
EI and Aboriginal and Torres Strait Islander research	22
<i>Overarching Issues Common to both ERA and EI</i>	22
Frequency of ERA and EI	22
Streamlining and simplifying ERA and EI	23
Utilising technological advances and pre-existing data sources	25

¹ The Australasian Council of Deans of Arts, Social Sciences and Humanities (DASSH) is the authoritative agency on research, teaching and learning for the Humanities, Arts and Social Sciences (HASS) in Australian and New Zealand universities. DASSH supports those within these institutions who have responsibility for the governance and management of research, teaching and engagement in HASS disciplines. DASSH also supports those who aspire to these positions through a Network of Associate and Deputy Deans.

Excellence in Research for Australia

ERA policy

Value of ERA

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.**

A moderate amount.

DASSH members supplied a variety of responses to this question, ranging from 'a small amount' to 'a large amount.' On balance, 'a moderate amount' is the most representative response.

Despite improvements in the last, more refined and demanding round, the ERA still 'misses' too much quality research by relying on too narrow a range of measures. For the HASS disciplines, the reliance on metrics still disadvantages researchers who, by other, more appropriate measures of their work, are doing quality research.

Those arguing for 'a small amount' note that the exercise is extremely expensive, both to the government and to the universities. The process is highly prescribed, takes inordinate amounts of time from researchers in universities and in government to collect data in a way that does not add value nor provide rich data, over and above what universities already know from their own understandings informed by international rankings. There is a sense that ERA runs too frequently, and it is impossible to compare across disciplines because of the use of citation analyses in predominantly the STEM disciplines and peer review in predominantly the HASS disciplines.

Some members argue that, despite its shortcomings, the ERA has encouraged Australian researchers to prioritise quality over quantity, which is a good thing.

Members generally doubt that the evaluations are well understood by Government, industry and community. There does not appear to be much interest in research quality in policymaking, only in short term instrumentalist research. In any case, many members feel the year to year changes and trajectories implied by these are confusing and often artefacts of the process, so the exercise does not provide assurance of maintenance of growth in quality.

- 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.**

A moderate amount.

Members appreciate this goal, but most also consider the process flawed. There is concern about consistency in assessment of submissions and about institutions 'gaming' the system, for example through 'strategic hires' or false claims of eligibility in certain disciplines.

There are some concerns that the exercise still allows for institutions to game the systems i.e. rearranging of publications into specific FOR codes to ensure the highest possible outcome. Those universities with sufficient funding are still able to improve their ERA rankings by bringing in high-ranking researchers from other universities. Overall, this trading of scholars between universities does not lead to any overall improvement in research outcomes across Australia.

As the exercise assesses research according to how it is classified in the ANZSRC, interdisciplinary research outputs are often obscured, as is research in fields which do not appear in the ANZSRC. It remains to be seen as to whether the revision of the ANZSRC this year will fully address these issues.

Some regional members noted that there is no counterweight for the difficulty small disciplines in smaller regional universities face in making the threshold, irrespective of the quality of research quality of the outputs. This puts in question the accuracy of the 'national' stocktake.

c. Identify excellence across the full spectrum of research performance.

A small amount.

ERA is based on publication outputs and funding. This is only a small component of research performance and omits other important research performance metrics.

There is an enormous disadvantage in areas of peer review compared to citation-based analysis, as peer review relies on subjective analysis. Our members have found that many fields are reliant on narrow pools of reviewers, which casts doubts on whether meaningful conclusions can be drawn from the exercises. The outcomes of peer reviews can be uneven and harsh across disciplines, especially in HASS, where the majority of peer review is undertaken.

Within HASS, there are substantial problems in HCA assessment where bibliometrics are not useful and peer review process is too internalised, lacks realistic international peer comparison and can be overly critical. There are problems with analysing quality in creative research, which are partly to do with the small national arts system and lack of opportunity to disseminate works, but also with on-going and uneven development of a culture of research in arts practice.

Sustainable research excellence is achieved by complex systems of mentoring, researcher development and generous collegiality. Unlike some other research evaluation processes that are attempting to craft 'responsible metrics', ERA's metrics are based in competitiveness and create active barriers to collegiate, generous, mentoring.

d. Identify emerging research areas and opportunities for further development.

A small amount.

Emerging research areas and opportunities are generally scored poorly in the ERA exercise so rather than being viewed as emerging, they are viewed as at risk. For example, while a ranking of 2 or 3 might be considered as identifying an opportunity for further development, because ERA is primarily reputational, it can also encourage institutions to deprioritise that area and try to fall below minimum threshold in future. Furthermore, because it is retrospective, gaming the submissions process by universities rewards existing success and managing coding so as to push some FoRs below submission threshold to hide weaknesses.

The rankings are not fine-grained, and the end result is a blunt tool containing no expert advice or direction for either institutions or individual researchers in terms of what the best avenue for further development might be.

Finally, the degree of gaming that ERA enables and promotes suggests that its outcomes are not a reliable reflection of the research landscape, its health or sustainability, and its retrospective nature limits the utility of insights about potential.

e. Allow for comparisons of research in Australia, nationally and internationally, for all discipline areas.

A moderate amount.

The end result is a set of rankings, so there is capacity for comparison of institutions in Australia. The fairly minimal metadata made available in relation to each institution's submission, though, means that the

national comparison is blunt and has little useful information for institutions wishing to learn from others and improve their performance.

The ERA exercise has become more about the strength of the university rather than research comparisons and the data has very little use outside of Australia. Lack of international peer assessors in HCA means that robust international comparison is generally not possible. Assessors from DASSH universities have indicated that Australian rankings are unrealistically low on international comparisons and believe that this is largely as a result of over critical and competitive expert assessor in small fields. Furthermore, schemes internationally work at different scales and to different logics (see, for example, the UK and New Zealand schemes). The comparability between schemes is limited.

Some members also doubt ERA's capacity to allow for international comparisons at all. No international institutions are specified in the rankings and 'world standard' is a nominal standard presumably developed internally by the ARC and the REC committees, but which is not broken down in great detail for researchers or institutions as stakeholders. Others still are asking the ARC to consider 'responsible metrics'² which are being investigated for use in the next UK REF and promote the use of indicators and underlying data infrastructure that recognises the true variety of research impacts and measures of quality.

As in the response to (b) of this question, there is a sense that ERA comparisons disadvantage regional universities which lack the economies of scale of larger metropolitan universities with greater capacity. This is thrown into relief with smaller disciplines.

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

a. If you disagreed with the previous statement, what should the primary purpose of ERA be going forward?

Neither agree nor disagree.

The objectives are appropriate, but the underlying methodology is problematic.

At universities, stakeholders are those within the university, our colleagues in other universities and potential partners (including industry). The purpose and objectives of ERA are not always appropriate to the needs of all stakeholders.

Under objective (a), industry and business are the same thing as organised societies such as NGOs, NPOs, charities, cooperatives have been condensed into 'wider community' which is not appropriate. We recommend the use of the word "Industry" only and to define that very broadly. The alternative would to say "government, industry, civil society organisations and the public".

Stakeholders are only given access to the four-digit FOR code rankings (1-5) but this does not provide the fine-grained analysis of the narrative reports. The researchers themselves, as primary stakeholders, may gain from ERA, because excellence is rewarded, but those in need of more support may not be well served by ERA. The ERA system does not offer a means to improve weaker results and rewarding the successful leads to equity issues across universities. Individual researchers are not given the chance to read the narrative reports, which are confidential, so the ERA exercise has little meaning for researchers (other than those involved in leadership) and does not provide a strong incentive to improve performance.

Objective (e), if it is to be retained, requires a significant change to the evaluation process. ERA evaluation provides an adequate national comparison, but almost no level of international comparison. International comparison would require benchmarking against other selected countries and significantly increased levels of international peer-review (at least 50% of all reviewers).

² See <https://www.responsiblemetrics.org/about/>

Q3.3 What impacts has ERA had on:

- a. the Australian university research sector as a whole**
- b. individual universities**
- c. researchers**
- d. Other?**

A – Australian University Research Sector.

If nothing else, ERA has had a substantial impact on the sector because of its function as a mandatory, national research reporting exercise. Some members feel this impact is limited to creating more work for university staff already burdened with compliance and administrative responsibilities, while producing little in the way of tangible outcomes. While collaboration within an institution may be fostered through ERA, it does pose significant implications for fostering national (and international) collaborations.

On the positive side, it has brought the question of research excellence to the forefront for the sector, and for both individual scholars and institutions. Its inclusion of creative works as NTROs has ensured that creative research practitioners in universities have seen that their work is valued by government, and it has given creative researchers a formal language with which to describe the research value of their work (albeit a language required by the ERA so its function is self-referential; creative researchers have become better at formulating the value of their work for the purposes of ERA).

B – Individual Universities.

Those contributing to the ERA exercises frequently see the activity as yet another overlay of work that must be undertaken, a factor which is especially significant at smaller, less well-resourced institutions. This is work that takes researchers away from their core business of conducting excellent research. ERA reports are useful for exposing areas of strength and weaknesses and provide an interesting analytic to compare FoR coding and performance across organisational units. The top-down logic of the FoR coding tends to show the customary and often arbitrary nature of internal organisation and highlights instances where internal organisation impedes collaboration.

Some institutions have adopted a strategy of investing in multidisciplinary and interdisciplinary initiatives in order to maximise the strengths of disciplines that perform best in the ERA and in EI. This has had the effect of focussing priorities on key disciplines. There are a number of benefits (particularly in EI), but the outcomes are not necessarily a reflection of quality but rather critical mass in a university without large capacity.

C – Researchers.

Researchers who are engaged in the ERA exercise spend six months of a year devoted to ERA rather than their research, while researchers not engaged in the exercise are generally unaware of the processes. The outcomes are beneficial for benchmarking individual and group performance, but also problematic in that it is not sensitive to the ways individual researchers have very different opportunities to achieve in research. It is not clear that the benefits of ERA can justify the input required, with researchers describing it as draining, alienating and demotivating.

Q3.4 How do you use ERA outcomes?

There is a sense in some institutions that ERA has become an exercise in strategy rather than in meeting the needs of stakeholders such as government, politicians, industry, small businesses, students, fellow researchers, and the Australian community. In many institutions the exercise is not associated with any forward-looking objectives.

Universities use ERA outcomes to determine where disciplines are excelling well above world standard and above world standard (compared to other universities with expertise in the same discipline) and identify

where disciplines are not meeting these standards and attempt to understand why. Generally, the reasons are diminishing staff, change of staff or research concentration in the area, rather than anything to do with researcher performance. Nevertheless, a low ranking (3 or lower) can be a problem and is demoralising for those working in these disciplines. This may be especially true in HASS disciplines which are seeing significant cuts in smaller institutions this year, creating problems of lack of coverage, depth in some areas.

For research directors, ERA may be used to understand relative performance within a discipline nationally or in comparison to other disciplines within the university. The data has been useful in explaining the relative success of smaller disciplines within universities on a national scale, when the scale of that discipline's research looks trivial when compared to larger disciplines within the university.

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

Disagree.

Our members note the high workload associated with the ERA collection and most doubt whether the outcomes are proportionate to the time and effort of research and administrative staff. Some members indicated that staff in their faculties are not aware of how the results of ERA are used beyond the initial reporting of results. Researchers have described ERA as a retrospective process without a clear rationale or specific benefits for individual researchers and the organisation. They suggest greater clarity is provided about how the ERA is used.

As a rankings exercise, the chief value of ERA outcomes is an outward-facing measure of a discipline's profile. Some members believe ERA is 'loosely valuable' in promoting performance in particular discipline areas, and as an independent, peer-affirmed measure of quality that preferences quality over quantity. This is countered by others who report that ERA reports have directly contributed to their institutions narrowing their research foci (e.g. continuing to support only disciplines in which they already rank highly).

In terms of the external responses to ERA, members report that the disciplines performing well in ERA have not experienced any notable increase in undergraduate or HDR students or research income either from grants or industry partnerships. Likewise, disciplines that perform less well have had not seen a decrease in students or funding.

Q3.6 Do you have any suggestions for enhancing ERA's value to you/your organisation?

Yes – see below.

- Consider how to capture the ways research in particular disciplines contributes to industry development, as well as opportunities to develop world-class capability. This needs to be considered concurrently to ensure necessary research capacity at the national level can be parleyed into research at a global level.
- Provide evidence of particular disciplines using both qualitative and quantitative data e.g. evaluations of the quality of key articles/books; qualitative evaluations alongside citations.
- Excerpts from academics about how their research informs their teaching – particularly of large cohorts of students; courses relevant to professional practice (at both undergraduate and postgraduate levels; e.g. teacher education; guidance & counselling); courses that contribute to academics' engagement in community.
- Reduce cost and burden on universities by shifting to a system in which the ARC collects the bulk of data in an automated way.
- Address the rating score inequities between citation metric disciplines and peer-review disciplines.
- Consider a shift towards responsible metrics or more ethical metrics so that ERA is better able to capture the components that underlie the production and sustenance of research excellence.

ERA methodology

ERA methodology at a glance

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

Disagree.

Our members recommend there be a review of the methodology across the disciplines and consideration of the implementation of a blended approach using both citation analysis and peer review in certain disciplines. This would likely provide a far more accurate assessment of disciplines in the humanities and social sciences, for instance, where both books and journal articles are common outputs.

The peer review process remains essential for many disciplines but requires improvement. Many disciplines too small a pool of expert assessors and there is a problematic shortage of international assessors given the criteria of the ranking. Peer reviewers need to be trained (including to ensure they have a thorough understanding of the discipline as a whole rather than just their particular field) and careful scrutiny of the peer reviewers is necessary to ensure that any biases are removed.

Q3.8 What are the strengths of the overall methodology?

Members appreciate the ERA methodology's robust definition of 'research' and emphasis of quality over quantity, and some believe this has had a positive impact on research activity.

Q3.9 What are the weaknesses of the overall methodology?

- **Thresholds:** Universities gaming the thresholds to exclude problem areas. There is a problem with small disciplines that are important but where the whole nation together has difficulty meeting the threshold. The focus on where the majority of outputs lie simply does not recognise the complexity of 'excellence'.
- **Income Data:** Use of income data at thresholds in ranking does not reflect the scale of income in small humanities disciplines.
- **Peer Review:** The quality and the calibre of Peer Review assessments is problematic. Volunteer reviewers are often be early career or inexperienced scholars, while more experienced scholars (with a more thorough knowledge of their disciplines) do not because they have little incentive to do so beyond altruism. As a result, the peer review is of limited value to the ranking panel in many cases. Further, there are very few international reviewers. If properly done, process should include training of reviewers, at least 50% of whom should be sourced internationally to ensure outputs meet the "international standard".
- **Journal Rankings:** The ranking of journals has resulted in an undervaluing of smaller and emerging sub-disciplines and Australian publications as well as disciplines that don't place emphasis on citations. Some of the ERA requirements are counterintuitive in terms of the EI outcomes. For example, if Australian stakeholders are more likely to read targeted Australian research publications, then emphasising international publications may reduce national impact and benefit to Australia.
- There is a mismatch between a six-year period being measured and a three-year frequency of assessment. Frequency and time scale should be aligned, with a five-year period being optimal (i.e. one ERA every five years assessing five years of data).

Citation analysis methodology

Q3.10 The citation analysis methodology for evaluating the quality of research is appropriate.

Strongly Disagree.

In the case of many humanities disciplines, citation analysis is of little value in determining research excellence. There is presently no useful method of capturing citations particularly in books and chapters. Within disciplines, fields of research can be small and often independent, so the scales are not comparable.

The speed of research means that lags in citations of relatively novel research can be a decade. Truly significant research often reaches acknowledgement in citation several decades after publication.

Members considered the citation thresholds to be too low. It is too easy to attain rankings of 4 and 5 in the disciplines assessed using citation analysis.

Q3.11 Does the discipline-specific approach for evaluating research quality (citation analysis or peer review for specific disciplines) continue to enable robust and comparable evaluation across all disciplines?

While members agree that citation analysis is entirely inappropriate for many HASS disciplines, there is also a strong consensus that there are weaknesses in the peer review methodology that must be addressed.

Given that metrics have now become more established in academia, even in HASS disciplines, it may be time to enhance purely 'peer review' methods of review with 'peer review informed by metrics' methods for reviewing outputs in HASS disciplines. Assessors could then review the influence of a researcher or output within an academic discipline that is independent of EI measures of non-academic influence. The review methods need not be either/or, although, for most HASS disciplines this would require taking stock of book/book chapters as well as journal citations.

Q3.12 What are the strengths of the citation analysis methodology?

Citation analysis avoids introducing the personal biases observed in peer review assessment.

Q3.13 What are the weaknesses of the citation analysis methodology?

Citation analysis can tend to favour established, highly renowned researchers over even some of the most excellent early career researchers.

The databases on used for citation analysis are notoriously variable in their capture of citations for different kinds of outputs. Reliance on any single database will inevitably reflect disciplines unevenly.

Q3.14 Can the citation analysis methodology be modified to improve the evaluation process while still adhering to the ERA Indicator Principles?

a. If you answered 'Yes', please describe how the methodology could be improved.

N/A.

Peer review methodology

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

Disagree.

Peer review is an acknowledgement that citation practices in HASS are not the same as in STEM, which is indeed the case; and it also allows for submission of NTROs which likewise cannot be picked up by citation indices, so it's appropriate in that sense. The requirement to have a relatively high proportion of all research available for peer review is a hard and useful test, and a good way to understand the quality of the whole submission.

Nevertheless, members are not convinced that the way peer review works in practice is comparable to how citation counts work in the STEM disciplines. Peer review assessments will always be, to a greater or lesser extent, subject to biases (conscious, unconscious or structural), either within the criteria provided to reviewers by the ARC or within reviewers and REC members themselves.

The evaluation of NTROs in past ERA iterations has been uneven, with a lack of clarity on how to assess this part of the 1904 submission. There have been very few '5' scores in the 1904 category (just one in 2018), with a unit appearing to need the disciplinary equivalent of a Nobel Prize winner on staff or to have had

made serious investment into more quantitative music disciplines (e.g., empirically-based research). Neither of these scenarios provide indicators of intrinsic research quality per se, but rather speak to assessors looking for external indicators of quality rather than peer reviewing in the traditional sense.

Other concerns expressed by members include the lack of international benchmarking in peer review assessments, the 'gaming' of the system through targeted recruitment of adjuncts, and the apparent tendency of the methodology to reinforce existing structures rather than identifying and encouraging opportunities in new areas of research.

Q3.16 What are the strengths of the peer review methodology?

Peer-review, by definition, focuses on the quality and academic merit of the research and its scope, theory, epistemology, and methodology. It allows for nuanced assessments which recognise the variability in what constitutes a 'world standard' across disciplines. There are also opportunities for institutions and researchers to articulate the intentions behind what they do, and particularly in the case of NTROs, to tell the story of the creative work, its provenance and its impact.

Peer review is a better option for disciplines that cannot rely on citations as a measure of impact. It allows for a more 'contextual' and considered assessment of quality of publications, such as standing of publisher, the ways the scholarship is reaching into new areas, using novel methodologies etc.

Q3.17 What are the weaknesses of the peer review methodology?

The assessment of quality is not subject to the rigour of comparison outside of the review process. Only reviewers and editors carry out the evaluation.

There is no training for assessors, nor are assessors provided with any feedback after the process. There needs to be more professional development on how to conduct the review from a discipline expert level. Currently, the inconsistencies in the quality and calibre of the peer reviews make the assessments problematic. The lack of training, combined with the significant work undertaken by reviewers, can result in what some have termed 'volatile' assessments.

In practice, there are almost no international Peer Reviewers, despite the stated intentions. The process should involve at least 50% international reviewers to ensure outputs meet the "international standard."

With regard to NTROs, the 30-megabyte limit on research output samples is restrictive, as it requires splitting video outputs into multiple files for example, or just submitting one brief excerpt of a larger work. The result is that different institutions deal with this limitation in different ways, meaning that reviewers can be presented with a wide range of volumes and types of evidence for a given discipline.

Q3.18 Can the peer review methodology be modified to improve the evaluation process while still adhering to the ERA Indicator Principles?

a. If you answered 'Yes', please describe how the peer review methodology could be improved.

Yes.

Suggestions from members include:

- Providing comprehensive training and guidance to reviewers, including a definition of "international standard" and advice on how to assess in relation to this term. This could include international benchmarks as examples for different disciplines.
- Including some 'assessment of the assessments' to reduce the disadvantage in rankings experienced by peer reviewed disciplines in comparison to disciplines assessed through citation analysis.
- Implementing a mixed model using both qualitative and quantitative indicators for most disciplines.
- Discipline specific committees to advise the ARC. HCA panel to provide examples of World Standard to stop criteria standards creeping upward.

- Including a high proportion of international expert assessors in each discipline code.
- A study of the scale of small disciplines nationally with a view to adjusting down the thresholds for submission where appropriate.
- Rather than 30%, a set number of items were to be peer reviewed, as in the REF in the UK, where each researcher submits four items, for instance. This might make for a more even playing field and ameliorate some of the effects of scale currently driving ERA results.
- Reduce the overall volume to be assessed by individual reviewers.
- Allow for scale in NTROs, as many institutions already do, differentiating between a 'standard' NTRO worth the equivalent of a journal article, and a 'major' NTRO equivalent to a research monograph.
- Compensate reviewers for their time to ensure they are not sacrificing their own research or working unpaid hours in service of the exercise.
- The inclusion of applied activities in assessments may be appropriate in the case of some HASS disciplines, but the applied measures listed in the in the Consultation Paper look remote, for example, to the discipline of Music. Nothing 'cultural' in nature is included in the list.

Contextual indicators

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

Agree.

The three contextual indicators relevant to the HASS disciplines remain important to ERA in providing points of reference for assessing qualitative indicators such as peer review and quantitative bibliometrics. The quality of the outputs should be the key focus, above and beyond achieving a certain critical mass. however, the volume threshold already ensures that the only units with sufficient mass are reviewed.

Some members were concerned that reviewers are asked not to consider productivity, even though this provides important context in the same way that 'trajectory' does.

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

Disagree.

This has become much less useful because of changes in publishing practice, in particularly those caused by Open Access publishing. The ARC's own requirements, if taken seriously, mean that outcomes are unable to be published with prominent academic presses. Much of the important new research in some creative fields is effectively self-published online. The only indicators of quality are the reputations of the people and institutions involved.

Given that this contextual indicator "has virtually no effect on the rating given to a unit of evaluation", it is unclear what value or function it has in the review process. A better test of quality may be to focus on a few chosen items for peer review and to replace this indicator with a citation profile to reflect the unit's level of influence in the discipline or field. If it remains, it should also be applied to NTROs as well as traditional outputs.

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

Agree.

In the case of Category 1 funding, they reflect a rigorous process of peer review of the grant. Category 2 and 3 funding also are useful to represent the demand for the research from industry and community. Research income data allows for useful comparison of disciplines for research management purposes that goes beyond the assessment process.

Members cautioned, however, that research income is a means to an end. It is means of providing resources to make research activity that leads to quality outcomes possible. It should not be evaluated on its own as a measure of quality without consideration of the outputs it generates.

There were also some concerns about the use of the census date in relation to this indicator, noting that it is subject to manipulation through hiring practices. Using the average FTE over the duration of the review period may help to address this issue.

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

a. Patents.

N/A.

b. Research commercialisation income.

N/A.

c. Registered designs.

N/A.

d. Plant breeder's rights.

N/A.

e. NHMRC endorsed guidelines.

N/A.

ERA rating scale

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

Neither agree nor disagree.

Five bands is generally considered appropriate, but the 'world standard' nomenclature is a nebulous measure that is not defined in the ERA supporting documents. Either the ranking scale should be adjusted to one more like that used for EI, or 'world standard' should be more explicitly defined.

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?

a. If you answered, 'Yes', please explain how the rating scale can be modified to identify excellence.

Not necessarily.

The large proportion of units scoring at or above world standard likely reflects that Australia is generally above world standard.

An argument could be made to expand to a 6- or 7- point scale to allow for a more refined set of indicators of where various units sit in comparison to their competitors.

ERA low-volume threshold

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

Disagree.

There need to be more discipline specific thresholds based on a national count of research outputs by discipline. Some small disciplines are so small that only one or two universities can reach the threshold, which indicates the threshold needs to be lower. Excluding small disciplines leads to good research being intentionally mis-coded so that it can be included, and the activities of whole small disciplines disappearing. Recent changes in the FOR codes for Creative disciplines (splitting into specific disciplines on performance, music etc) will result in smaller universities being unable to meet the threshold despite the high level of quality work.

In the case of some larger disciplines, the thresholds are too low, allowing relatively large amounts of poor research to be hidden in below-threshold units of evaluation.

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

There needs to be more discipline specific thresholds based on a national count of research outputs by discipline. This cannot be based on past ERA data as the universities have cross coded or recorded outputs

to game the system at the expense of distorting the description of the research. It may be better for the ARC, rather than institutions, to allocate codes.

The low volume threshold could also be adjusted when there are 2 or more codes that are related, and both have a moderate amount of outputs. This is particularly relevant for creative disciplines that have recently had the theory and practice codes separated.

ERA staff census date

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

A combination of the two. Relying on census dates only has resulted in institutions ‘gaming’ the process.

If it can only be one or the other, the by-line approach is preferable. It is more a genuine reflection of staff’s connection with an institution and helps to “reduce incentives to engage staff merely for the purpose of claiming all their research outputs within the reference period”.

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

Census dates promote practices not conducive to a thriving research sector, e.g. ‘poaching’ of personnel and the hiring of adjuncts. A university is allowed to claim the outputs of honorary staff and staff who have not been given ongoing employment providing they were employed during the census period.

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

Maybe.

The by-line approach may be better overall given the evident gaming of the system around census dates, however there is scope for a by-line approach to be gamed too. There will also be complexities around which university deserves to be acknowledged in the by-line as it will remain difficult to judge if more researchers end up being ‘floating’ researchers.

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

Some journals/books only allow a by-line from one institution when some staff have multiple institutional affiliations, and some books and NTROs do not carry a by-line at all.

There is scope for a by-line approach to be gamed too, such that a department can claim the works of a given scholar but does not in reality have their presence and input into mentoring, and collegial support of research development.

ERA interdisciplinary research and new topics

Q3.31 ERA adequately captures and evaluates interdisciplinary research.

a. If you disagreed with the previous statement, how could interdisciplinary research best be accommodated?

Disagree.

The process encourages disciplines to silo their outputs into the largest disciplines for maximum results. The assessment by discipline effectively disables ERA’s capacity to effectively capture interdisciplinary research and its value. The exercise actually creates a sense of competition across codes further de-incentivizing interdisciplinary research.

Interdisciplinary work was not well reflected by the previous set of FOR codes, and thus was difficult to truly capture. It is unclear what effect the revised ANZSRC will have on interdisciplinary research reporting in ERA, but it seems unlikely that there will often be sufficient outputs to meet the low-volume thresholds in these codes.

ERA and Indigenous research

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

a. Two-digit?

N/A.

b. Four-digit?

N/A.

Q3.33 In ERA, the best approach for evaluating Indigenous Studies is:

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.

D – Other.

Indigenous Studies most often entails very different ways of comprehending, understanding and being in the world from those deployed in other disciplines, most especially with regard to Australian Indigenous communities. The ERA approach is philosophically opposed at a fundamental level to the belief systems of the First Australians. A better approach to evaluating Indigenous Studies would be to formulate an evaluation process that grew from the principles of Indigenous practice and should be undertaken in close collaboration with Aboriginal and Torres Strait Islander scholars.

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

If the ERA is to be used to publicise research, then we should aim to publicise Indigenous Studies more and it is currently likely that Indigenous Studies will often fall below the current threshold. Implementing a new system tailored to the specific character of Indigenous Studies research would allow more of it to be assessed and publicised.

ERA process

Collection of ERA data

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

Neither agree nor disagree.

An annual collection may be advantageous as it would involve a small increase in work in an ongoing capacity rather than a very major piece of work few years. However, it is difficult to assess without more detail about the processes involved in an annual collection are required to provide a more definitive answer to this question. This also assumes the continuation of the current practice of assessing all the research produced, data collection requirements would be very different were the ARC to adopt a different approach, such as the REF in the UK.

Q3.36 What would be the advantages and/or disadvantages of an annual data collection?

It would depend on what reporting the annual ERA required. The exercise would need to be radically simplified and/or automated if it was to be done annually. Members strongly support an automated data collection (e.g. using ORCID) by the ARC.

Publication of ERA data

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

Yes.

This would increase the ability of audiences to understand the context for an institution's outcome in a given discipline and thus improve the ability to compare institutions at a national level.

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

- a. What would be the advantages?
- b. What would be the disadvantages?

Agree.

Members agree that ERA should be more transparent. Showing more metadata would provide greater context for its outcomes. In the case of publishing the assignments of disciplines to research outputs by institutions, this would expose and perhaps go some way to preventing intentional gaming of the exercise.

The report would result in a huge amount of data, though, and could be controversial in the sector, though this may not necessarily be a long-term disadvantage.

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

Members provided the following suggestions:

- The volume of each institution's submission in each discipline, and its relation to FTE staffing in that discipline.
- Total time used by each institution to prepare ERA submission and participate in reviews/panels
- Estimated national cost
- Feedback by the panel on each FOR to each institution
- One narrative example of excellence per discipline code per university to avoid compounding the inequity between the large and small universities.

Engagement and Impact Assessment

EI Overview

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

- a. **encourage greater collaboration between universities and research end-users, such as industry, by assessing engagement and impact?**

A small amount.

University researchers who collaborate with industry already had strong engagement with a pathway to impact. The EI exercise simply encouraged better documentation of the activities, and there are likely better incentives for researchers to increase their collaboration with end-users rather than a stocktaking exercise.

EI works towards ends (applied research, practitioner-oriented outcomes) that are not the same as those of ERA (scholarly excellence). These different drivers demand different kinds of performances, both of which are currently being asked simultaneously of all 'evaluation units.'

- b. **provide clarity to the Government and the Australian public about how their investments in university research translate into tangible benefits beyond academia?**

A moderate amount.

The Impact Case Studies and Engagement narratives provided excellent examples to the government and Australian public about how their investments in university research translate into tangible benefits, but this was just a small snapshot of an inordinate amount of activity.

c. identify institutional processes and infrastructure that enable research engagement?

A moderate amount.

EI did encourage and assist universities to examine their institutional processes and infrastructure to enable research engagement, though it is unclear as to whether this translated into improving those processes in a systematic manner.

d. promote greater support for the translation of research impact within institutions for the benefit of Australia beyond academia?

A moderate amount.

More resources are being devoted to the promotion of research outcomes, and more researchers are thinking about and engaging with a plan for research impact. EI provided greater transparency in terms of research impact and there are more conversations in the sector about research impact and the translation of research. However, EI has not been the only contributing factor in these changes. Some members argue that making engagement and impact a priority in grant applications is a more effective way to create change, along with recognition in internal university incentive systems such as promotion.

e. identify the ways in which institutions currently translate research into impact?

A moderate amount.

The approach did focus some greater institutional attention on what they had been doing in relation to research translation and may have encouraged greater investment of resources in research translation, at least prior to the pandemic.

Q4.2 The EI objectives are appropriate for the future needs of its stakeholders.

Neither agree nor disagree.

Within universities, members agree that EI has been a valuable exercise, but are sceptical of its reach outside the university sector. It is not clear that the objectives and definitions of EI are understood by potential stakeholders.

Members do agree that it has provided a useful prompt for universities to attend to impactful research. There have been many anecdotal cases where the exercise made visible the work of researchers that is having a profound impact for good, but which may not be rewarded under current university protocols.

Q4.3 What impact has EI had on:

- a. the Australian university sector as a whole?**
- b. Individual universities.**
- c. researchers.**
- d. other sectors outside of academia?**

A – Some disciplines are unclear of the benefit of EI and see it as a significant diversion of resources in to an unwieldy, time and resource expensive exercise that produced unclear results. Other disciplines felt that EI has made the impact of university research more visible and recognised previously undervalued research, particularly those in non-traditional formats.

B – The exercise has encouraged stronger conversations about research impact and a greater push for engagement with industry, government, and communities. It has allowed universities to better understand, articulate and value the significance of research in terms of its real-world influence.

C – Greater awareness of pathways to impact for their research.

For some researchers the focus on non-academic outcomes is welcome, while others struggle to find examples of engagement and impact resulting from their work, but this has been a valuable exercise for all. However, in a climate where there are increasing time pressures and growing expectations that researchers do everything, clarity around what really matters, and the timing of activities is needed.

D – Very little, if any.

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

Members reported that the outcomes were used to applaud disciplines that excelled in the exercise, and the Impact Stories were used to publicly showcase the university's external engagement.

With regard to NTROs, the exercise has been useful as it acknowledges that creative works, such as stories or artworks, disseminate information and knowledge in ways different to a scholarly article but which do have a high value. In this sense being able to note impact and engagement gives value to non-traditional research outcomes.

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

Neither agree nor disagree.

The EI process promotes forward planning for impact, which has value in encouraging researchers to connect with research end users, but the process of documenting impact case studies is extremely labour intensive and arguably of limited intrinsic value.

Q4.6 How else could EI outcomes be used?

They provide a narrative for showcasing the outstanding research impact that researchers have across the university sector.

EI definitions

Q4.7 The current Engagement definition is appropriate.

a. If you don't agree, what are your suggested amendments to the Engagement definition?

Neither agree nor disagree.

The definitions themselves required considerable reflection and explanation to allow researchers to understand and consider their own practices through these prisms. A clearer and useful definition would considerably improve the process.

Q4.8 The current Impact definition is appropriate.

a. If you don't agree, what are your suggested amendments to the Impact definition?

Neither agree nor disagree.

The definitions themselves required considerable reflection and explanation to allow researchers to understand and consider their own practices through these prisms. A clearer and useful definition would considerably improve the process.

Q4.9 The current end-user definition is appropriate.

a. If you don't agree, what are your suggested amendments to the end-user definition?

b. Are there any end-user categories excluded in the current definition of research end-user that you think should be included?

Neither agree nor disagree.

The definition is broadly acceptable, but the exclusion of impact within the higher education sector is problematic for research in education, human resources and similar disciplines.

Q4.10 Are there other key terms that need to be formally defined?

N/A.

EI methodology

Unit of assessment

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

Yes.

There was strong consensus against having more case studies among members.

Members report that two-digit codes work well, with several disciplines cohering into a strong whole supported by data and narratives from fields within and beyond HASS.

Assessment using the two-digit codes provides scope for capturing the breadth of activity to accommodate the diverse portfolios of kinds of work that scholars do, from the purely theoretical to the more applied/immediately impactful.

For many smaller units this is the only meaningful level at which to make such an assessment

Q4.12 Are there other ways to classify units of assessment in EI, for example, SEO codes?

Yes.
Using SEO codes would be feasible; however, it is useful to examine results from non-academic impact (EI) and academic quality (ERA) assessments alongside one another to see the full breadth of research activity for a code.

Selectiveness of EI

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

a. How many and why?

The same or fewer units of assessment.

Fewer units of assessment (say, 6-10) would allow a reasonable representation that might produce more sustainable resource demands on universities in their efforts to comply. Some members proposed an opt-in policy for each institution since not all units will have interesting impacts to report in every round.

EI low-volume threshold

Q4.14 The EI low-volume threshold should continue to be based on the number of research outputs submitted for ERA.

a. If you disagree, how should eligibility for assessment in EI be determined?

Neither agree nor disagree.

Not necessarily. Members note that some units generate substantive impacts and rich engagement that are not reflected in peer-reviewed research outputs. An alternative would be that universities could be allowed to nominate their own 'best performers'.

Q4.15 The low volume threshold is set at the appropriate level.

Neither agree nor disagree.

Engagement indicators

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

Disagree.

Members note apparent deficiencies in scope to include the types of impact generated by some HASS disciplines, particularly creative and cultural disciplines. These indicators could be optional, and the emphasis instead placed on the narrative that contextualises the nature and the extent of the engagement.

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

Disagree.

There is a strong consensus among members for the EI exercise should include for in-kind contributions as they are an important indicator of engagement with many sectors that do not have the capacity to make significant financial contributions. This is especially evident in HASS but is also true of many disciplines that

work with the not-for-profit sector. The Linkage Program shows there are already ways of measuring and recording in-kind contributions.

Relying on cash support disregards and undervalues, for example, engagement with vulnerable communities and values research purely in commercial terms. If 'cash support from research end-users' is included as an indicator, the surrounding narratives should make clear why it is appropriate.

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

Disagree.

Relying on commercial income disregards and undervalues, for example, engagement with vulnerable communities and values research purely in commercial terms. If commercial income is included as an indicator, the surrounding narratives should make clear why it is appropriate.

Q4.19 Are there additional metrics that would be appropriate across many or all disciplines?

Yes.

Philanthropic income is currently not included in measures of engagement because it is regarded as insignificant. However, inclusion of philanthropy would encourage behavioural change and drive a greater engagement with that sector. Philanthropy is critically important in the translation of pure research to commercial outcomes, more so than industry engagement (in the US, industry engagement is around 6% and philanthropy around 30%).

There is a strong consensus among members for the EI exercise to account for in-kind contributions as they are an important indicator of engagement, especially in HASS. Linkage projects show there are already ways of measuring and recording in-kind contributions.

Engagement with Indigenous communities and groups could be included as a metric to promote and encourage stronger emphasis on work of this nature.

Members have recommended the ARC undertake research to identify the forms of 'responsible' and 'ethical' metrics being considered in other national systems.

There is also an argument from some that metrics and indicators should be optional, and the emphasis should be placed on the narrative that should contextualize the nature and the extent of the engagement.

Q4.20 Are there alternative metrics that would be appropriate across many or all disciplines?

Yes.

- Philanthropic Income
- In-kind Income
- Engagement with communities, including Indigenous communities
- Book sales
- Digital media metrics
- Policy advice
- Advisory body roles and consultations
- Research collaborations and associations, including with community groups
- Partnerships with end-users
- 'Responsible' / 'ethical' metrics

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

Yes.

Some members were critical of metrics which rely on income only as a measure of engagement or impact, suggesting, for example, that such metrics discriminate against disciplines that work with vulnerable communities.

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

Neither agree nor disagree.

Members felt that as this was not a common or suitable practice in HASS disciplines it would provide an unreasonable burden that may prompt less than best practice in supervision.

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

- a. Patents.
- b. Research commercialisation income.
- c. Registered designs.
- d. Plant breeder's rights.
- e. NHMRC endorsed guidelines.

N/A.

These metrics may be appropriate for some disciplines but are irrelevant for most HASS disciplines.

Engagement narrative

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

- a. If you disagree, what alternative approach could be used to replace the narrative?

Agree.

It provides the flexibility to allow engagement to be captured in all its appropriate fit-for-purpose formats. It would be challenging to develop metrics that would sufficiently address research engagement with end-users. The narrative approach allows for the communication and drawing together of conceptual frameworks around the research.

Several members requested that future iterations include clearer directions to universities, for example, guidelines about what should be included in the engagement narrative and a sense of what is seen to be the 'correct' or 'ideal' balance of Category 1 vs. Category 2-3 research.

Q4.25 One engagement submission per broad discipline is sufficient for capturing the research engagement within that discipline.

Agree.

There was strong consensus against having more case studies, particularly if moving to more than one submission means doubling the amount of work.

A balance needs to be struck between the number of cases, their development and evidence and the substantial costs incurred in gathering evidence, constructing narratives and developing cases. The impost for the 2018 round was substantial. Given the resource constraints in the sector of the coming years, it is not conceivable that universities will have the resources available to participate without significant implications for the core functions of research and teaching.

Q4.26 The engagement narrative needs to be longer.

Neither agree nor disagree.

A balance needs to be struck between the number of cases, their development and evidence and the substantial costs incurred in gathering evidence, constructing narratives and developing cases. The impost

for the 2018 round was substantial. Given the resource constraints in the sector of the coming years, it is not conceivable that universities will have the resources available to participate without significant implications for the core functions of research and teaching.

Q4.27 Additional evidence is needed within the narrative.

a. If you agreed, what evidence should be provided?

Neither agree nor disagree.

A balance needs to be struck between the number of cases, their development and evidence and the substantial costs incurred in gathering evidence, constructing narratives and developing cases. The impost for the 2018 round was substantial. Given the resource constraints in the sector of the coming years, it is not conceivable that universities will have the resources available to participate without significant implications for the core functions of research and teaching.

Impact narrative

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

a. If you disagree, what alternative approach could be used to replace the narrative?

Agree.

The narrative approach provides the flexibility to allow engagement to be captured in all its appropriate fit-for-purpose format and for units to contextualise the nature and the extent of their impact.

It should also be recognised that the impact of ideas is hard to measure. An idea can snowball across groups, communities and time. It can grow, change and shift in meaning. Therefore, measuring impact is difficult and inexact.

A balance needs to be struck between the number of cases, their development and evidence and the substantial costs incurred in gathering evidence, constructing narratives and developing cases. The impost for the 2018 round was substantial. Given the resource constraints in the sector of the coming years, it is not conceivable that universities will have the resources available to participate without significant implications for the core functions of research and teaching.

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

Agree.

There can be small units with powerful impacts. Two suggestions came from members, either to have a number of studies proportionate to the outputs or it could be made optional for universities to decide.

Q4.30 The impact narrative needs to be longer.

Disagree.

It should be possible to make a case in a minimal space and reduce the burden on assessors.

Q4.31 There is a need for additional evidence to be provided within the narrative.

a. If yes, what evidence should be provided?

Neither agree nor disagree.

Narratives backed by additional evidence would be made stronger, but this should be taken into account in their evaluation. To reduce administrative burden, the provision of evidence need not be made into a requirement for all such narratives. Minimally, affidavits should be provided from end-users about the link between research and its impact, detailing both qualitative and quantitative indications of its significance and strength.

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

a. **If you answered 'yes' to the previous question, please name and describe the quantitative indicator/s, and the disciplines for which they are relevant.**

Yes.

It depends on the nature of the impact. For example;

- Number of systems improved
- Number of people affected
- Number of organizations involved in uptake

Approach to impact Narrative

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

a. **If you disagree, what alternative approach could be used to replace the narrative?**

Agree.

The narrative provides the flexibility for units to contextualize the nature and the extent of their impact. It allows the institution to describe the complex and multiple drivers for impact that metrics cannot explain.

Q4.34 One approach to impact narrative per broad discipline is sufficient for capturing the activities within that discipline.

Agree.

There can be small units with powerful impacts. Two suggestions came from members, either to have a number of studies proportionate to the outputs or it could be made optional for universities to decide.

Q4.35 The approach to impact narrative needs to be longer.

Strongly disagree.

It should be possible to make a case in a minimal space and reduce the burden on assessors.

Q4.36 There is a need for additional evidence to be provided.

Strongly disagree.

Evidence can be described in the narrative.

The impost of generating evidence is substantial. Raising the demand for evidence will incur significant costs on universities in what is already an expensive exercise.

Q4.37 Would there be benefit in combining engagement and approach to impact?

Yes.

Engagement and Approach to Impact are intertwined and their separation in the assessment is unclear, as the lack of definitional clarity illustrates. Combining the two elements would also reduce administrative burden.

El rating scales

Q4.38 The engagement rating scale is suitable for assessing research engagement.

Disagree.

In the 2018 round, there was a lack of clarity around how to interpret the rating scale.

Q4.39 The descriptors for the engagement rating scale are suitable.

Disagree.

In the 2018 round, there was a lack of clarity around how to interpret the rating scale.

Q4.40 The impact rating scale is suitable for assessing impact.

Disagree.

In the 2018 round, there was a lack of clarity around how to interpret the rating scale.

Q4.41 The descriptors for the impact rating scale are suitable.

Disagree.

In the 2018 round, there was a lack of clarity around how to interpret the rating scale. The scales in the pilot exercise appeared to make more sense.

Q4.42 The approach to impact rating scale is suitable for assessing approach to impact.

Neither agree nor disagree.

Q4.43 The descriptions for the approach to impact rating scale are suitable.

Neither agree nor disagree.

EI interdisciplinary research

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

Yes.

Interdisciplinary research is likely to become more prominent and it may be impactful. Including it provides a solution to the limitations of working within FOR codes.

EI and Aboriginal and Torres Strait Islander research

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?

No.

Returns from universities that are below the threshold will be difficult to compare with universities that are above. The method would have to be changed first.

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

Yes.

Engagement is a core principal of research in Indigenous Studies and should be included.

Overarching Issues Common to both ERA and EI

Frequency of ERA and EI

Q5.1 How often should ERA occur?

Every five years.

The exercise is costly in both time and money and should not be conducted more often than every five years. The biggest problem members reported in relation to these exercises is the time and effort required to participate in them; time they take away from research, teaching and engagement. A shorter timeframe also would be a disadvantage to early career scholars and disincentivise departments to hire more people.

Allowing longer timeframes (e.g. five years or more) is advantageous in that it allows time to reflect on the outcomes of the previous round and some opportunity to institute changes in practices and emphases (and possibly to measure these). Research excellence and research cultures takes years to develop and mature.

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?

A longer time allows for greater reflection. Research cycles for projects are often longer than three years, and it takes yet more time to gather evidence of sustained impact.

Members also suggest longer timeframes would lessen the propensity of institutions to game the exercises. It would reduce the artificial 'churn' of staff associated with ERA-driven recruitment ('head-hunting').

The absence of a three-year cycle would encourage collegiality and 'high-risk, high-return' research ideas to flourish. Shorter timeframes narrow the willingness to enable development and the appetite to risk failure.

Q5.3 How often should the EI assessment occur?

Every five years.

Engagement and, especially, impact take years to develop and mature. A five-year timeframe would give more time for research to reach end-users between assessments.

Members would like timeframes to be consistent between ERA and EI.

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?

As well as maintaining consistency with ERA, it would allow adequate time for trusting relations for engagement to develop between researchers and partners and thereby make the path to meaningful impact much less circuitous and chance riven.

Streamlining and simplifying ERA and EI

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

a. What would be the advantages and/or disadvantages?

Neither agree nor disagree.

There was considerable disagreement among members in response to this question.

In large part, it would depend on how onerous the exercises are in future. If the ARC collects and sorts the review data and only relies on institutions to assist with peer and panel review, then the two can be combined. If the existing method is retained, then combining the assessments would be too onerous to perform in a single year.

It would also depend on whether the purpose and intent of the EI exercise is to complement ERA. There needs to be more clarity around the actual intent of the EI exercise. These assessments measure quite different things that should not be confused.

Advantages:

- Integrating ERA and EI may provide a more complete picture of a unit's research trajectory.
- Integration may reduce the overall administrative burden of the exercises, but this is by no means guaranteed.

Disadvantages:

- Assessing engagement and impact alongside research excellence may lead to an erosion or devaluing of blue-sky research.
- The diversity of forms of excellence in research may be overlooked.

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

Yes.

Any mechanism to reduce the reporting burden on universities may be supported in principle, but without details it is difficult to assess risks and benefits e.g. although shifting to automation of FoR coding for

publications may streamline some reporting, universities would still need the capacity to change pre-assigned codes where necessary to reflect the actual content.

Suggestions include:

- ARC collecting all data using an automated system
- Automation of FoR coding
- Some disciplines currently assessed via peer review could include at least partial assessment using citation metrics e.g. Divisions 35, 38, 39, 43, 44 (formerly 13, 14, 15, 16, 21)
- Only including C1 outputs that carry DOIs to exclude many 'vanity' and 'predatory' publications
- ARC pre-filling the forms
- Adopting a different approach to assessment in which not all publications are submitted but only a small sample (e.g. the method used by the UK's REF)

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

Members suggested the ARC could use ORCID, Google Scholar, Clarivate, but note that all have weaknesses. A combination of data sources may produce better results.

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

a. Are there efficiencies that could be introduced?

Time-consuming elements:

- The Peer and Panel Review components
- Selection of Peer Review cross section.
- Gathering and assessing publications/NTROs, especially for peer review.
- Assembly of data, which requires substantial time contributions from research offices, finance officers, individual researchers, Associate Deans (Research) and Discipline Champions.
- Checking data for comprehensiveness
- Following on 'missing' publications
- Attending meetings to assess modelled outputs and negotiate on the distribution of publications across cognate fields.
- Training staff on 'performing for ERA'
- Providing labour to 'mock assessment' exercises.

Efficiencies:

- Don't overcomplicate. Any money required to undertake these quality control checks takes away from doing the research in the first place.
- Devise a system that is not vulnerable to gaming.
- Lower thresholds to reduce the amount of writing required.

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

a. Are there efficiencies that could be introduced?

Time-consuming elements:

- Imagining potential forms of evidence
- Collecting and collating evidence
- Translating evidence in narratives across multiple 'contenders'
- Participating in mock assessments
- Selecting and refining submissions

Efficiencies:

- Don't over complicate. Any money required to undertake these quality control checks takes away from doing the research in the first place.
- Refine and clarify definitions.

- Remove reliance on metrics that are evidently biased in terms of capturing only engagement and impact that is primarily or only expressible in financial terms.

Utilising technological advances and pre-existing data sources

Q5.10 ORCID IDs should be mandatory for ERA?

a. What are the advantages and/or disadvantages?

Agree.

Wherever possible, ERA should use existing publication monitoring systems to reduce the administrative burden on universities, though some members note that ORCID has significant weaknesses in relation to books, non-English language scholarship and NTROs.

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

a. What are the advantages and/or disadvantages?

Agree.

Wherever possible, ERA should use existing publication monitoring systems to reduce the administrative burden on universities, though some members note that ORCID has significant weaknesses in relation to books, non-English language scholarship and NTROs. Relying on ORCID alone would not be appropriate.

Some members have reported that the reliance on ORCID during grant proposal preparation in 2019 made the process more time consuming than in previous years when publications were input manually.

Q5.12 DOIs should be mandatory for ERA.

a. What are the advantages or disadvantages?

Agree.

Members support this proposal as it would make the reporting and assessment processes more efficient but add the caveat that it should only apply for outputs where this is possible. Outputs that are not digitized must be exempt and not excluded altogether.

Q5.13 Are there new ways to collect data to reduce the cost and burden to universities of participating in ERA and EI whilst maintaining the robustness of the ERA and EI process?

a. What are the advantages and/or disadvantages?

Yes.

Use existing monitoring systems where possible, e.g. ORCID, Clarivate, Google Scholar, ARC RMS, DOIs.