# RESEARCH ASSESSMENT IN AUSTRALIA: JOURNAL RANKING, RESEARCH CLASSIFICATION AND RATINGS

**Gaby Haddow**
School of Media, Creative Arts and Social Inquiry, Curtin University, Australia
G.Haddow@curtin.edu.au
https://orcid.org/0000-0002-2038-6436

*Abstract*
Australia introduced a new research assessment model in 2010 after many years of assessing research through purely quantitative measures. Excellence in Research for Australia (ERA) assesses research quality across the entire higher education sector and involves extensive data collection and highly detailed submissions for universities. This chapter traces the history of research assessment in Australia before discussing three key components of the ERA model - journal ranking, research classification and assessment ratings. It focuses the discussion on how these components of the ERA have impacted the social sciences and how the social sciences have fared under the model. Drawing on national and international commentary and research, this critical analysis of research assessment in Australia seeks to interrogate the current model and inform future developments in research assessment.

*Keywords*
Research assessment; social sciences; Australia; journal ranking; research classification; assessment ratings

**CHAPTER**

*Introduction*
Research assessment in Australia has evolved from a simple quantitative system that counted research outputs to a complex model that aims to assess quality, engagement and impact. This chapter will chart that history, focusing on the Excellence in Research for Australia (ERA) assessment model that was first implemented in 2010 with the aim to assess research quality.

The ERA is a national research assessment exercise in which the entire university sector participates. It is a cumbersome model that requires extensive data collection by institutions, a substantial number of unpaid assessors and ongoing review by the coordinating agency, the Australian Research Council (ARC). Given the amount of time and resourcing needed for an ERA round, it is remarkable that no direct funding is attached to the model. Instead, institutions receive ratings of research quality for over 170 fields that make up the research classification scheme.

A ranked journal list was used to support quality assessment in the first ERA round, despite disquiet across the sector and particularly the social sciences. Although the ranked list was not carried across to the 2012 ERA, the use of journal ranking has lodged in the Australian scholarly communication psyche. It seems that institutions and discipline areas remain wedded to using some form of ranked journal list.

ERA ratings are assigned on a scale of one to five, with the highest rating indicating research quality 'well above world standard'. The process and decision-making that leads to an ERA rating is the subject of increasing criticism. Different assessment techniques are applied for social sciences and

humanities (peer review) and the sciences (citation analysis) and over the life of the ERA substantial variation in ratings has been observed, to the detriment of the peer review disciplines. Another concern related to ratings is the scant information and feedback provided to the sector about how decisions to assign ratings are made, with commentators remarking on a lack of transparency.

The ranked journal list, research classification and ratings are, or have been, key components of the ERA. The ranked journal list was discarded in 2012, a 2019 review of the research classification scheme resulted in a completely revised set of fields of research, and a review of the ERA is underway. It is timely then to consider how the social sciences have fared under this model. To do this, the chapter draws on national and international commentary and research, as well as the responses in interviews conducted by the author in 2017 with 51 social sciences researchers in Australia (Haddow & Hammarfelt, 2019). Firstly, a background to research assessment in Australia provides the context for its development and ultimately the ERA model. The chapter then critiques the creation and use of ranked journal lists, the research classification scheme and the ERA ratings process and outcomes. A brief discussion of the relatively recent Engagement and Impact (EI) exercise concludes the body of the chapter. Quotes from the interviews are used to illustrate how individual social scientists have experienced journal ranking and research classification in the Australian research assessment environment.


### *Background*
It is now 30 years since the Australian government introduced a system to assess university research performance (Butler, 2003). The changes to research assessment that have occurred in Australia over this period sparked increased awareness and interest amongst the academic community and led to the publication of research findings that remain important to research evaluation scholars to this day.

During the early 1990s, Paul Bourke (1994) and Linda Butler examined Australia's research outputs to identify publication formats and international visibility through indexing by, what was then called, the Institute of Scientific Information (ISI). They found over 60 per cent of Australian journal articles were published in ISI-indexed journals, noted the diversity of publication forms produced by Australian researchers and remarked on the inadequacy of a single indicator to evaluate research. The strong representation of Australian publications in international journals at that time was not unique to Bourke and Butler's work (Butler, 2017, p. 920) and it demonstrated the outwards focus of Australian researchers; a perspective that is reflected in research assessment in Australia still.

Australia's early performance-based research funding system, the Research Quantum (RQ), distributed funding to institutions based on research grant income, initially, with publication numbers and research degree completions added later to the funding model (Butler, 2017). The diversity of publishing in the social sciences and humanities was recognised by the RQ model, so that a book chapter attracted the same number of points as a journal article, and books were rewarded with five times the points available for a journal article. In a groundbreaking study published in 2003, Butler raised a significant concern about the Australian RQ model. The allocation of funding based on a simple count of publications had resulted in what she described as 'entirely predictable' outcomes; that was a substantial increase in publication numbers but at "the lower end of the impact scale" across all disciplines (Butler, 2003, p. 39). Butler observed that the funding model meant "there is little incentive to strive for publication in a prestigious journal" (p. 41). The lack of 'quality' or 'impact' criteria in the RQ model was addressed in two ways. Firstly, by the early 2000s the Australian government was exploring alternative research assessment models, particularly the Research Assessment Exercise (RAE) in the UK (Butler, 2017). Secondly, and much later, the publications count was removed from the RQ funding model.

For ten years, the Higher Education Research Data Collection (HERDC), which was the basis of the RQ model, has run in parallel with a new research assessment model, the ERA. With one exception, between 2011 and 2016, when the ERA results were used to inform an initiative called Sustainable Research Excellence leading to block grants to universities, no funding is associated with an ERA assessment (Woelert & McKenzie, 2018, p. 193). Universities were therefore reliant on the funding that flowed from the HERDC submissions for three indicators - research income, publications and research degree completions until 2015. In the years that followed, the indicators for publications and research degree completions were removed from the HERDC submission criteria so that only research grant income remains in 2020. Research degree completions attract funding from different sources, the Research Training Program and the Research Support Program (Department of Education, Skills and Employment Australia, n.d.; Woelert & McKenzie, 2018).

*From the Research Quantum to the Research Quality Framework*
The Australian government announced a major change to how research would be assessed in May 2003 (Butler, 2008). Sir Gareth Roberts, who had reviewed the UK's RAE, was appointed to chair an Expert Advisory Group to develop the new model, called the Research Quality Framework (RQF). A preferred model was released in 2006 and its purpose was to fund research based on research quality and impact. The RQF followed similar lines to research assessment models in the UK, Hong Kong and New Zealand. Proposing a transparent process and aiming to reward research excellence, the RQF's criteria for assessment – quality and impact – related to the significance or merit of research in the academic community and the use or transfer of research outside the academic community, respectively (Haddow, 2007).

The proposed units of assessment in the RQF model were research groups of five or more academic staff working at lecturer level or above. An RQF submission would comprise an evidence portfolio covering the previous six-year period, and include: a statement about the group's focus, strategy and activities; publications overall and the four best for each group member; research income; and statement of impact. A research classification category and keywords could be selected by the submitting group to describe their field and these would guide the decision as to which of 13 expert panels would assess the submission. A five-point rating scale for quality and another for impact would be assigned to a unit of assessment. The quality scale ranged from 'research that does not meet the standard of recognised quality' to 'world leading research or research that is highly significant to Australia'. Groups rated at the lower end of the scale would not receive funding (Haddow, 2007, p. 27).

Metrics were "recommended to inform decision making" in the RQF assessment process (Butler, 2008, p. 86). Butler was appointed to chair the RQF's Quality Metrics Working Group (QMWG) and this group rejected the use of impact factors and citations without contextual information as assessment tools. However, three citation-based calculations were proposed with the recommendation that citations data collection was outsourced to an independent agency to guarantee consistency across submissions (Haddow, 2007, p. 30). The proposed use of citations for all disciplines raised concerns for editors, academics and scholarly communities. A study of social sciences journals indicated that a relatively low percentage of the 244 titles examined "attract sufficient citations to make such an assessment meaningful on the article level" (Haddow & Genoni, 2010, p. 483).

Another proposed 'quality' assessment tool was a list of ranked publication outlets. Presumably founded on Butler's 2003 research findings, the QMWG determined that journals and other research outlets would be ranked "to encourage researchers to publish in the most prestigious outlets for their discipline" (Butler, 2008, p. 90). The responsibility for ranking journals, conferences and

creative arts venues in four tiers, A*, A, B and C, was delegated to disciplinary groups. As a general guide, groups were advised to assign approximately 5 per cent of outlets as A*, 15 per cent as A, 30 per cent as B and the remainder as C (Butler, 2008). By 2006, several discipline areas including education, creative arts, and management had developed rankings for journals and conferences (Haddow, 2008). Of the many ranking exercises that took place, the Australian Business Deans Council (ABDC) (2019, p. 4) "inaugural version of the ABDC Journal Quality List" in 2008 is the most notable and enduring.

Late in 2007, Australia held a federal election which spelt the end of the RQF. As Hicks (2010 p. 33) wrote several years later: "This elaborate exercise was the result of an extensive consultation and design exercise, but was abandoned by a new government before implementation in favour of the ERA, which emphasises simplicity". As will be illustrated in this chapter, 'simplicity' was never achieved despite the new model focusing solely on quality.


### *Excellence in Research for Australia (ERA): 2008-2020*
The new government announced that the first ERA exercise would be conducted in 2010 (Box, 2010), and since then evaluations have taken place in 2012, 2015 and 2018 (Crowe & Prado, 2020). The objectives of the ERA are:
- Promoting Excellence: Rigorously assess research quality to promote pursuit of excellence across all fields and all types of research.
- Informing Decisions: Provide a rich and robust source of information on university research excellence and activity to inform and support the needs of university, industry, government and community stakeholders.
- Demonstrating Quality: Provide government and the public with evidence of the quality of research produced by Australia's universities.
- Enabling Comparisons: Allow for comparison between Australian universities against world stand for all discipline areas (Australian Research Council, 2021b).

The ERA model has been relatively stable over the past decade, with one exception - the ranked journals list.

The ERA is managed by the Australian Research Council (ARC, 2021b), the agency responsible for distributing national competitive research grants to non-medical/health sciences disciplines. An ERA assesses the previous six year's research outputs, the previous three year's research income, and applied and esteem factors for all research-active staff (over 76,000 in 2018) at 42 eligible institutions. Almost a third of the research workforce are associated with the social sciences and humanities.

Disciplines and researchers are classified with Field of Research (FoR) codes from the Australian and New Zealand Standard Research Classification (ANZSRC), administered by the Australian Bureau of Statistics (2008). The ANZSRC is a "hierarchical classification that has three levels, division (two digits), group (four digits), and field (six digits). … The two-digit codes consist of the collection of the related four-digit FoR codes" (Crowe & Prado, 2020, p. 2). Research quality is assessed at the two-digit and four-digit level for the ERA.

The documentation for an ERA submission is extensive (see 'Key Documents', Australian Research Council, 2021b). In essence, institutions must submit the details of all publications and non-traditional outputs over the assessment period. A full version of the outputs must also be made accessible to assessors. Eligible outputs for the social sciences and humanities are: books, book chapters, journal articles, conference publications, and non-traditional outputs such as creative

works. Each output must be assigned at least one and up to three four-digit FoR codes. With the exception of journal articles, an institution can select the most appropriate FoR code for an output.

Journal articles are different. As noted above, journal ranking activities were being undertaken in Australia as early as 2006 and a draft ERA journal list was finalised in 2008. Rankings of A*, A, B or C were assigned to the journal titles in the first iteration of the list, which comprised over 20,000 titles. Each title was assigned at least one and up to three FoR codes by the scholarly academies and disciplinary groups that participated in the ranking exercise (Genoni & Haddow, 2009). In practice this means that an article is assigned the same classification as the journal in which it is published. A small percentage of journals, approximately 3.5 per cent, are assigned a multidisciplinary (MD) classification (Haddow, 2015). Some flexibility is available in FoR code assignment. Institutions can argue that over 66 per cent of an article's content is outside the publishing journal's assigned FoR codes, however there is no data to indicate what proportion of article submissions take advantage of this option. Only journals included in the ERA Journal List are eligible outputs for an ERA assessment.

While institutions are able to select the most relevant FoR codes for books, book chapters, conference papers and non-traditional outputs, it is not without its challenges in terms of classification. This will be expanded on in the Fields of Research section below. Institutions are also required to refer to ERA lists of eligible publishers and conferences, which creates further limitations to the submission of research outputs.

Certain thresholds of outputs are required for assessment at the four-digit and/or two-digit FoR code level, described as a Unit of Evaluation (UoE). If a threshold is not met at the four-digit level those outputs will contribute to an assessment at the two-digit level. The final rating for UoEs is decided by panels of experts, called Research Evaluation Committees (REC). There are eight RECs and each is responsible for assessing a group of disciplines, such as 'Education and Human Society'. REC chairs and members are dominated by Australian academics with only five members from non-Australian universities in a cohort of 150 (Australian Research Council, 2018b).

Research outputs are assessed for different FoR codes by peer review or metrics, determined by the ARC (2021b, see the Discipline Matrix). With very few exceptions, science disciplines are evaluated using a variety of citation-based metrics. Social sciences and humanities are evaluated by peer review. The RECs are assisted in this mammoth task by researchers who nominate as an ERA peer reviewer; a role that involves no financial reward. Institutions must nominate 'a sample' of 30% of their outputs for peer review and non-traditional research outputs require an accompanying research statement that describes the research component of the work.

To this point the processes involved in an ERA submission are detailed but relatively clear. However, context and transparency are lacking for both the submission process, in which "there is little scope … for those who are subjected to assessment to provide any sort of contextualizing information in addition to the reported outputs" and the ratings, which have no "contextualizing information about the assessment scores" (Woelert & Yates, 2014, p. 181). The ratings assigned for a UoE are simply:

| | |
|---|---|
| 5 | Well above world standard |
| 4 | Above world standard |
| 3 | At world standard |
| 2 | Below world standard |
| 1 | Well below world standard |
| NA | Not assessed due to low volume |
| NR | Not rated |

'Not rated' has been used when a REC judges that an institution's submitted outputs are not properly aligned with the FoR code under which they are submitted.

The lack of context referred to (Woelert & Yates, 2014) could be alleviated by greater transparency around the ratings process. ERA submission data is not released as a data set after an assessment round, which leads to comments, such as the following by former Deputy Vice-Chancellor (Research) at the University of Melbourne:

> The quantitative and qualitative details of the benchmarks used by the ARC and how they have changed since 2012 are not available for independent scrutiny. Furthermore, the ARC does not provide any relative commentary on the assessment of individual disciplines or universities, leaving the research community to speculate on the significance of the findings. (Larkins, 2019b, para. 7)

While the ARC has invited feedback about the (unranked) journal list, the ANZSRC and other components of the ERA over the last 12 years, there is a sense that this is a closed system. For example, revisions to the ERA Journal List are not made available prior to the ERA round in which it is being used (Australian Research Council, 2021). This means that researchers and institutions are, in many cases, 'working in the dark' as they make publication decisions.

Harking back to the ideals of the original RQF proposal, the ARC introduced a second assessment model to evaluate engagement and impact (EI) in 2018. The aim of the EI was to: "assess how well researchers are engaging with end-users of research, and shows how universities are translating their research into economic, social, environmental, cultural and other impacts" (Australian Research Council, 2021a). Only three ratings were used to assess engagement and impact: high, medium and low.

The outcomes of the 2018 ERA and EI assessments were released in 2019, followed by an announcement that both models would be reviewed (Australian Research Council, 2021b). Future exercises are scheduled for 2023 and 2024, respectively. Also reviewed in 2019-20 was the ANZSCR FoR codes (Australian Research Council, 2020), which has subsequently undergone major revisions (Australian Bureau of Statistics, 2020).

In the discussion that follows, three key components of the ERA model - journal ranking, research classification and assessment ratings - are addressed in more detail and in the context of the commentary and research surrounding them. As a newcomer to Australian research assessment, the Engagement and Impact model has attracted less attention. An overview of this aspect of the Australian evaluation landscape is included later in this chapter.

*Journal Ranking*
Numerous examples of journal ranking exercises exist in the scholarly literature and many of the resulting lists have not survived the scrutiny they received. In Australia, a list of ranked journals was intended as an integral assessment tool for the RQF and the ERA (Butler, 2008). The task of ranking according to the 'quality' of articles published in journals was delegated by the ARC (with scant guidelines) to the four Australian learned academies and disciplinary peak bodies. These peak bodies were disparate in their membership and their approach to the ranking exercise. For example, the Australian Political Studies Association organised panels to undertake the initial ranking using some citations data, followed by feedback from members. The Australian Association for Research in Education gave priority to the needs of practitioners and asked members to list their 10 best journals, from which a ranking was produced (Genoni & Haddow, 2009). Around the same time the UK's Association of Business Schools (ABS) had created a list of preferred journals (Mingers &

Willmott, 2013), which was closely followed by the release of the Australian Business Deans Council "Journal Quality List" in 2008 (2019).

A critique of journal ranking was published soon after several lists, including the European Reference Index for Humanities in 2007, the Agence d'Evaluation de la Recherche et de l'Enseignement Superieur in 2008 and Australia's ERA list, were released and subsequently withdrawn. The authors note: "Far from being neutral, journal rating activities always involve stakes which are both cognitive in relation to the categories produced, and political with regard to their effects" (Pontille & Torny, 2010, p. 348). Ferrara and Bonaccorsi (2016, p. 281), commenting on journal ranking in the social sciences and humanities, stated there are "serious problems of fragmentation and lack of consensus". The lack of neutrality, consensus and anomalies in the Australian journal ranking exercise was raised by several disciplines at the time (Haslam & Koval, 2010; Lamp & Fisher, 2010; Vanclay, 2011). Epitomising a concern raised in the social sciences, was an analysis of employment relations journals that found very few Australian journals ranked as A* or A (Young et al., 2011). Additional disquiet occurred when the final 2010 journal list was released, as Cooper and Poletti (2011, p. 63) noted: "The erratic re-rankings that occurred between the last draft version of the journal rankings and the 2010 finalised list (where journals went from A* to C, with some disappearing altogether) have left many researchers uncertain as to whether current rankings will apply in 2012". These proved to be prophetic words.

The widespread resistance to the ERA's ranked journal list likely contributed to its abandonment for the 2012 ERA. The Chief Executive of the ARC had another explanation: "some institutions used the journal ranks out of context, potentially harming the careers of young academics and those working in cross-disciplinary areas" (Sheil, 2014, p. S67). Sheil responded to questions about the ERA process at the time, saying "the sector would have to give up a little transparency along with the journal rankings" (Woodward, 2011, para. 10). However, universities did not 'give up' the ERA ranked journal list, as demonstrated five years later by an interviewee:
> *The journal rankings are out of date and discredited but I am still expected to use them.*

Some institutions have adopted different forms of ranked journal lists to encourage researchers to produce 'quality' publications. Web of Science, Scimago and particularly the ABDC list were noted in the interviews, with a researcher in a business field stating:
> *It's publish, publish, publish but really we're only interested in A and A* as set out in the list of publications from the Australian Business Deans Council.*

In-house lists have also been created:
> *Recently the university has come up with a list of so-called quality journals which we're all now supposed to get additional credit for publishing in.*

An issue that was argued strongly in Australia (Genoni & Haddow, 2009; Young et al., 2011) is the concern that journal ranking can impact on how the social sciences engage in their national community. The ongoing use of ranked journal lists means that this concern remains, highlighted in the comment:
> *If you were to start to focus on these rankings in particular … you would orient your scholarship away from more local concerns.*

With journal ranking so entrenched in the research mindset and performance management processes in Australia, it is difficult to imagine a research assessment model that could deter their use. However, one aspect of the ERA relating to journal publications could be reconsidered and that is the pre-assignment of FoR codes.

*Field of Research Codes*

Research classification is a necessary component of research assessment in order to identify strengths, weaknesses and trends in the body of research being evaluated. However, the assignment of a classification system to individual journal titles, as is the case for the ERA, is uncommon (Goldfinch & Yamamoto, 2012).

The ANZSRC Field of Research (FoR) codes were assigned to journal titles during the journal ranking activities in Australia. While the rankings were removed from the journal list, the FoR codes remained for the more-than 20,000 titles deemed eligible journals for assessment. The 2008 release of the ANZSRC has been used in each ERA round (Australian Bureau of Statistics, 2008).

The assignment of FoR codes to journal tiles has received less attention than journal ranking, although Bourke and Butler (1998, p. 718) undertook research several years prior to the ERA's implementation that concluded "the most useful approach . . . are Field of Research (FOR) and subfield classifications based on journal sets".  A later study (Archambault et al., 2011, p. 69) noted concern about journal-level classification, stating "subject delimitations based on journal classifications are likely to contain articles that have a weak relation with the target subject". In an Australian study of music education journals only 50 per cent of FoR code allocations made by authors to their articles aligned with the codes assigned to the publishing journals (Bennett et al., 2011, p. 92). Another issue raised and relevant to the Australian model, is the inconsistency of assigning codes so that some journals have one code only, while others are assigned two or three (Archambault et al., 2011). This was explored for journals assigned a single FoR code, using visualization tools to assess the extent of content (drawn from abstracts) that lay outside the publishing journal's assigned FoR code (Haddow & Noyons, 2013). The study found that a single FoR code was an inadequate representation of journal content.

Because ERA ratings reflect the research quality and productivity of organisational units and the researchers associated with them, there is a strong motivation to publish in journals that are assigned corresponding FoR codes. Concerns about the impact of journal-assigned FoR codes were pointed out in a report by the National Tertiary Education Union, with one member commenting: "For every journal that they publish in outside of our department, it's not helping us" (Kwok, 2013, p. 35). An interviewee confirmed this dilemma:

> *if it includes [name of field] code … then that's fine.  But if it doesn't then I'd be wary of it because someone else will grab it for their ERA submission*

The extent of alignment of FoRs and organisational unit was tested on the journal publications of researchers at three social sciences and humanities 'schools'. Over 50 per cent of the publications were classified outside of the corresponding FoR codes for *education* and for *language, communication and culture* schools (Haddow, 2015, p. 335). A later study included books, book chapters and conference papers as well as journal articles for analysis, providing a truer representation of social sciences (Guns et al., 2018). With echoes of the earlier research, this exploration of cognitive and organisational classification found that some social sciences fields matched more closely across the two classification types than others, such as 'Communication studies'. In general, however, the researchers note that social sciences "publications are much more dispersed across different knowledge domains" than publications in the humanities (Guns et al., 2018, p. 1106).

Interdisciplinarity is perceived as important for knowledge production and innovation and the Australian government has repeatedly encouraged interdisciplinary approaches in its policy documents (Woelert & Millar, 2013). In practice, the ERA is counter to those stated aims, as Woelert and Millar (2013, p. 764) remark: "ERA appears to encourage strategic behaviours on the part of

universities and researchers that aim to hide rather than highlight the interdisciplinary nature of their research". It is a concern felt by individual social scientists, illustrated in the comment:

> There's definitely pressure though the ERA to publish in particular disciplinary journals, so journals that highly rate within a single discipline, rather than my work [which] is quite multi-disciplinary.

In some fields the number of eligible journals assigned an FoR code has the potential to influence publication choices, particularly those in the business disciplines that use the ABDC rankings:

> In the 1501 code there are very few journals. I don't think there are any A* ranked journals that are Australian.

For researchers who work in a field with limited 'eligible' publication options there are several disadvantages. These include a potential for reduced national discourse in the field because the more highly ranked journals are international (as noted by the interviewee) and the possibility that highly ranked journals are unsuitable due to their scope (for example, a focus particular research methodologies). To which, researchers may respond by adjusting their own research activities.

An additional complication in the ERA classification and submission system is that universities are responsible for assigning FoR codes to research outputs (or proportions of pre-assigned codes to journals), which makes it possible to promote or 'hide' a research field (Knott, 2015). It is a flaw in the ERA model that led to the introduction of the 'Not rated' rating and is demonstrated in the comment:

> In the worst case, [it] can lead to work not being counted because it doesn't fit the category that the school is applying for.

A review of the ANZSRC was announced in 2019 (Australian Research Council, 2020) and disciplinary groups were encouraged to make submissions with suggested amendments. The revised codes were released in 2020 with marked differences from the 2008 version. A major improvement is in the representation of Indigenous Studies, now with its own four-digit code. But it is at the six-digit level that most of the changes have been made. While many of the changes to the ANZSRC are welcome, a vast amount of work will be required to map the new codes to titles on the ERA list, if that occurs following the ERA review.

Classification systems are limiting, regardless of the care with which they are designed. It is the nature and purpose of classification to provide boundaries around categories, which in the case of the FoR codes relates to the subject content of research. The ANZSRC is probably unnecessarily granular, but it is the way the classification scheme is applied in the ERA that creates the issues described above.

*ERA ratings*
As noted above, the ratings given to UoEs are expressed in terms of international standards of quality and the ERA rating scale of 1 to 5 is unaltered since its introduction (see Australian Research Council, 2019, *ERA Reports*). This consistency should ensure that simple comparisons to identify trends can be made between one ERA exercise and another. However, the way in which final ratings are arrived at lacks transparency and institutions are provided with no information that contextualises the ratings assigned to UoEs (Woelert & Yates, 2014). That is, the ARC's use of the phrase 'world standard' comes without a definition or benchmark. In effect this suggests that individual assessors are left to devise their own measure of 'world standard' and those assessed are given no indication as to why research outputs do or do not achieve that standard. A clear example is the calculations/decisions applied to assess the contribution of a UoE at the four-digit level to its hierarchical two-digit level UoE (Crowe & Prado, 2020). For instance, in the 2018 ERA report for

'Education' (UoE 13), two institutions received different ratings for four-digit UoEs, yet both received the same two-digit rating. Other cases show the same rating for all four-digit UoEs at two institutions, but one institution is rated higher at the two-digit level. The table below presents an example from 'Education' (FoR 13).

| UoE | 13 | 1301 | 1302 | 1303 | 1399 |
|---|---|---|---|---|---|
| Institution A | 3 | 3 | 3 | 3 | n/a |
| Institution B | 3 | 2 | 2 | 3 | n/a |

Table 26.1 Example of ERA ratings

Concerns about the ERA ratings, and indeed the need to conduct an ERA at all, emerged after the first 2010 round (Marsh et al., 2012). These first ratings reflected the strengths of Australia's research-intensive universities, a collective known as the Group of Eight (Go8). Marsh et al. (2012, p. 86) reported that 91.3 per cent of the UoEs submitted by these universities were assessed at the highest rating. Focusing particularly on the social sciences and humanities, Larkins (2019a, p. 4) also noted the higher ratings for Go8 universities "in every ERA round".

Larkins (2019a) and Crowe and Prado (2020) also refer to 'ratings creep' as a potential issue for the ERA. There has been a minor increase in the number of UoEs assessed over the life of the ERA, from 642 to 677, but the proportion of UoEs assessed as above world standard has increased from 34.3 per cent to 56.6 per cent in the same period (Larkins, 2019a, p. 2). The apparent improvement in the quality of Australian research has Larkins (2019c, p. 2) asking: "Is this a real increase in excellence or have other factors influenced the outcomes?" Of concern to the social sciences is that ratings creep is associated with the way different disciplines are assessed. In the results of ERA 2018:

> 80 per cent of research in science, technology, engineering and mathematics was above or well above world standard, up from 43 per cent in 2012. The proportion of humanities, arts and social sciences research receiving similar appraisals increased just 8 percentage points to 35 per cent. (Ross, 2019, para. 3)

Larkins (2019c, p. 3) also asks "why the science-related and the humanities and social science trends are so different?" According to Ross (2019) these differences are not evident in the UK's research assessment exercise, which uses peer review for all discipline areas. The implication being that assessment by citation analysis results in higher ratings than peer review.

*Engagement and impact*
Assessment of research engagement and impact has had a patchy history in Australia: included in the proposed RQF, dropped for the ERA and now under review after one full round of the Engagement and Impact (EI) exercise in 2018. It is notable that despite the introduction of research impact assessment in the UK's Research Excellence Framework (REF) in 2014, Australia chose a broader remit by adding research engagement to its assessment model. Gunn and Mintrom (2018) view the inclusion of engagement favourably, in that it serves to disambiguate the two activities. Certainly, providing evidence of engagement, defined as "the interaction between researchers and research end-users outside of academia for the mutually beneficial transfer of knowledge, technologies, methods or resources" in the *EI 2018 Assessment Handbook* (Australian Research Council, 2018a, p. 6), is a simpler task than proving impact. Research impact is defined by the ARC as "the contribution that research makes to the economy, society, environment or culture, beyond the contribution to academic research" and the difficulties involved in demonstrating impact are numerous. Challenges exist in showing unique and direct links from the research to the impact outcome and the time lag between research and impact (Khazragui & Hudson, 2015). Moreover,

measuring impact has frequently relied on economic indicators, which are often unsuitable for the social sciences (Muhonen et al., 2020).

In the 2018 EI exercise, five assessment panels were established to assign ratings of Low, Medium and High to submissions at the two-digit FoR code level. Each assessment panel was responsible for a wide range of disciplines and, while research end-user members were included, the panels were primarily comprised of academic researchers. Engagement was assessed using information in a statement (narrative) and quantitative indicators relating to research income in the form of grants, HERDC funding and commercialisation income. Impact assessment relied on qualitative 'Impact studies', with an option to include quantitative indicators if appropriate. Separate ratings for engagement and impact were reported. Slightly fewer submissions, Units of Assessment (UoA) at the two-digit level, were lodged for EI 2018 than the ERA of the same year. Perhaps surprisingly, a higher proportion of the Impact submissions were rated as High (43%) compared to Engagement (34%). Unlike the ERA data, the ARC subsequently published 'Engagement narratives' and 'Impact studies' that were rated as High.

Many of the social scientists who participated in the 2017 interviews described their use of social media and academic networking platforms to gain exposure and link with the community and scholars in their field. There was some awareness that research assessment was entering a new phase in Australia with the EI model being piloted, and this was regarded as positive by nearly half the interviewee sample, for example:

> And hopefully we might move towards social impact, because ultimately I think it's a lot more important. A bit of research can change people's lives.

However, it was also clear that much of the research communication and outreach activities that the social scientists described were more aligned to engagement than impact.


### Conclusion

Several years before the introduction of the ERA, Laudel and Gläser (2006, p. 294) noted the tension between assessment and the "complex nature of communication in science". It is an issue that the ERA model was designed to avoid by using peer review to assess social sciences and humanities. However, a comparison of the ERA ratings suggests that the Australian model has been unsuccessful in resolving that tension and science fields have the upper hand in the Australian research assessment environment. The longer-term implications for social sciences and humanities are not encouraging:

> The three ERA rounds have demonstrated very clearly that if a university aspires to increase its overall research standing, including international rankings, then an investment in [science-related] disciplines is more likely to provide a better dividend than investment in the [humanities and social sciences] disciplines. (Larkins, 2019a, p. 7)

The ERA model is inherently complex. It combines a list of eligible journal outputs with a research classification scheme, affecting the accuracy of disciplinary representation for journal articles and creating unnecessary delineation issues for interdisciplinary research and for organisational units. Moreover, the early introduction of a ranked journal list has left an indelible 'standard' for quality publications. Australian universities were criticised for using the ranked list "out of context" (Sheil, 2014), but according to Woelert and McKenzie (2018) institutions have responded to the ERA as expected in applying national criteria at the institutional level. This trickle-down effect (Aagaard, 2015) is not unique to Australia and is a phenomenon that has been discussed widely (Butler, 2003; De Rijcke et al., 2016; Ingwersen & Larsen, 2014; Ossenblok et al., 2012; Woelert & McKenzie, 2018).

A significant concern is that Australian social scientists will respond to their institutions' research strategies by shifting their research from a local and national focus in order to publish in highly ranked international journals.

The fourth principle of *The Leiden Manifesto* (Hicks et al., 2015) is: "Keep data collection and analytical processes open, transparent and simple". Although these principles were created for the use of metrics, the overarching aim is relevant to research assessment more generally. As the discussion above illustrates, from submission to ratings, the ERA falls well short of this principle. In addition, the anticipated simplicity of the ERA to which Hicks (2010) referred appears to have been lost and the value of the ERA is questionable. The current exercise requires substantial resourcing from universities, but produces outcomes that merely reinforce existing understandings of research strengths and does not lead to direct government funding. While acknowledging the benefits of the ERA, Larkins (2019c, p. 6) contends there are now "diminishing returns for universities" and argues for a "simplified methodology for data collection" in a future assessment model.

Research assessment systems in other countries have the potential to direct developments in Australia. The long-standing REF in the UK has been used to inform several national models, including those in Italy, Norway and Sweden, although some selected only certain components and others were never implemented (Sivertsen, 2017). (Jon Holm's chapter in this book provides an excellent discussion of the Norwegian Model, first implemented in the late 1990s.) However, there are significant differences between the different assessment systems, in their purpose, components and their methodologies. Similar to Australia, the Norwegian Model uses a defined publications list that also includes indicators of quality. Unlike Australia, the Norwegian publications database is carefully curated for accuracy and currency, and publications are weighted according to field norms. In Norway and the UK, bibliometric indicators are used to inform assessment rather than to form the basis of assessment as they do for Australian researchers in the sciences. As discussed, the use of these quantitative indicators in the ERA appears to have favoured ratings for those fields. An important difference is also evident in the units of evaluation. In the ERA, a highly hierarchical and relatively inflexible classification scheme is applied to assign researchers and research outputs to fields, whereas the REF and Norwegian Model use organisational classification through the assessment of research groups within an institution.

These and a host of other differences are evident across research assessment systems. Each system is presumably designed to capture data that will inform national research strategy and, in many cases, the allocation of funds. But funding allocation is not a stated objective of the ERA. In fact, other than identifying "emerging research areas and opportunities for further development" (Australian Research Council, 2021b) the ERA objectives are somewhat vague and have little substance in terms of outcomes. It is arguable whether the cost to the public purse and to the institutions being assessed is justifiable simply to stocktake, identify excellence, reassure the community that Australia's research is excellent, and enable comparisons with other nations.

<a> Implications for research assessment
The Australian higher education sector is now waiting on the results of the current ERA review. If research, commentary and feedback are heeded, the eligible journal list with its pre-assignment of FoR codes will be abandoned. These changes would create more flexibility in publication choices and support, rather than restrict, interdisciplinary research. As the Norwegian Model has indicated, a list of eligible publications can provide a useful guide for researchers, but the management of that list is critical.

When the RQF, which proposed using citation-based indicators for all fields, was replaced by the ERA, which uses citation-based assessment for science disciplines only, it felt like the social sciences

had won the day. Unfortunately, the retention of quantitative indicators for the sciences has created an imbalance in the ERA ratings, to the detriment of the social sciences. A revised ERA would be well-served to follow the lead of other established research assessment systems and impose peer review (informed by indicators if appropriate) to rate all disciplines. To manage the additional work associated with peer review for all fields, the ARC could consider discarding the current criteria for submission data, which is the inclusion of all outputs over the assessment period, and introduce a formula for expected outputs, such as the REF (Research Excellence Framework, 2020).

The Australian Research Council has taken a step towards greater transparency in the EI 2018 with the publication of submission data (Engagement narratives and Impact Studies), which may spell changes for the ERA. A revised model must address the previous lack of transparency to ensure that institutions and individuals can see value in the significant resourcing commitment for an ERA submission. In its current form, the ERA's underlying assessment processes and opacity in ratings means it raises more questions than informs about research excellence in Australia.

## References

Aagaard, K. (2015). How incentives trickle down: Local use of a national bibliometric indicator system. *Science and Public Policy, 42*(5), 725-737. https://doi.org/10.1093/scipol/scu087

Archambault, É., Beauchesne, O. H., & Caruso, J. (2011). Towards a multilingual, comprehensive and open scientific journal ontology. In B. Noyons, P. Ngulube, & J. Leta (Eds.), *Proceedings of the 13th International Conference of the International Society for Scientometrics and Informetrics* (pp. 66-77). ISSI. https://www.issi-society.org/proceedings/issi_2011/ISSI_2011_Proceedings_Vol1_09.pdf

Australian Bureau of Statistics. (2008). *Australian and New Zealand Standard Research Classification (ANZSCR), 2008*. https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1297.02008?OpenDocument

Australian Bureau of Statistics. (2020). *Australian and New Zealand Standard Research Classification (ANZSCR): A statistical classification used for the measurement and analysis of R&D in Australia and New Zealand.* https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/1297.0Main+Features12020?OpenDocument

Australian Business Deans Council. (2019). *Australian business deans council 2019 journal quality list review final report*. https://abdc.edu.au/wp-content/uploads/2020/03/abdc-2019-journal-quality-list-review-report-6-december-2019_2.pdf

Australian Research Council. (2018a). *EI 2018 assessment handbook*. https://www.arc.gov.au/file/10666/download?token=62kvwQjS

Australian Research Council. (2018b). *ERA 2018 research evaluation committees (RECs)*. https://www.arc.gov.au/excellence-research-australia/era-2018/excellence-research-australia/era-2018-research-evaluation-committee-recs

Australian Research Council. (2019). *ERA reports*. https://www.arc.gov.au/excellence-research-australia/era-reports

Australian Research Council. (2020). *ANZSRC review*. https://www.arc.gov.au/anzsrc-review

Australian Research Council. (2021a). *Engagement and impact assessment*. https://www.arc.gov.au/engagement-and-impact-assessment

Australian Research Council. (2021b). *Excellence in research for Australia*.
https://www.arc.gov.au/excellence-research-australia

Bennett, D., Genoni, P., & Haddow, G. (2011). FoR codes pendulum: Publishing choices within
Australian research assessment. *Australian Universities' Review, 53*(2), 88-98.
http://hdl.handle.net/20.500.11937/22871

Bourke, P. (1994). Quantitative research indicators: Citations treatment 'deficient'. *Campus Review,*
9.

Bourke, P., & Butler, L. (1998). Institutions and the map of science: Matching university departments
and fields of research. *Research Policy*, *26*(6), 711-718. https://doi.org/10.1016/S0048-
7333(97)00046-2

Box, S. (2010). Performance-based funding for public research in tertiary education institutions:
Country experiences. In *Performance-based funding for public research in tertiary education
institutions: Workshop proceedings* (pp. 85-126). OECD.
https://doi.org/10.1787/9789264094611-en

Butler, L. (2003). Modifying publication practices in response to funding formulas. *Research
Evaluation, 12*(1), 39-46. https://doi.org/10.3152/147154403781776780

Butler, L. (2008). Using a balanced approach to bibliometrics: Quantitative performance measures in
the Australian research quality framework. *Ethics in Science and Environment Politics, 8*(1)*,* 83-
92. https://doi.org/10.3354/esep00077

Butler, L. (2017). Response to van den Besselaar et al.: What happens when the Australian context is
misunderstood. *Journal of Informetrics, 11*(3), 919-922.
https://doi.org/10.1016/j.joi.2017.05.017

Cooper, S., & Poletti, A. (2011). The new ERA of journal ranking: The consequences of Australia's
fraught encounter with "quality". *Australian Universities' Review, 53*(1), 57-65.
https://www.learntechlib.org/p/54268/

Crowe, S. F., & Prado, C. (2020). Excellence in research in Australia: The soufflé keeps on rising.
*Australian Psychologist*, *55*(5), 468-487. https://doi.org/10.1111/ap.12453

De Rijcke, S., Wouters, P. F., Rushforth, A. D., Franssen, T. P., & Hammarfelt, B. (2016). Evaluation
practices and effects of indicator use - A literature review. *Research Evaluation*, *25*(2), 161-
169. https://doi.org/10.1093/reseval/rvv038

Department of Education, Skills and Employment Australia. (n.d.). *Higher education research data collection (HERDC)*. https://www.education.gov.au/higher-education-research-data-collection

Ferrara, A., & Bonaccorsi, A. (2016). How robust is journal rating in humanities and social sciences? Evidence from a large-scale, multi-method exercise. *Research Evaluation*, *25*(3), 279-291. https://doi.org/10.1093/reseval/rvv048

Genoni, P., & Haddow, G. (2009). ERA and the ranking of Australian humanities journals. *Australian Humanities Review, 46*, 5-24. https://press-files.anu.edu.au/downloads/press/p39271/pdf/book.pdf#page=9

Goldfinch, S., & Yamamoto, K. (2012). *Prometheus assessed? Research measurement, peer review and citation analysis*. Chandos.

Gunn, A., & Mintrom, M. (2018). Measuring research impact in Australia. *Australian Universities' Review, 60*(1), 9-15.

Guns, R., Sile, L., Eykens, J., Verleysen, F. T., & Engels, T. C. E. (2018). A comparison of cognitive and organizational classification of publications in the social sciences and humanities. *Scientometrics, 116*(2), 1093-1111. https://doi.org/10.1007/s11192-018-2775-x

Haddow, G. (2007). Academic libraries and the research quality framework. *Australian Academic & Research Libraries, 38*(1), 26-39. https://doi.org/10.1080/00048623.2007.10721265

Haddow, G. (2008). Quality Australian journals in the humanities and social sciences. *Australian Academic & Research Libraries, 39*(2), 79-91. https://doi.org/10.1080/00048623.2008.10721334

Haddow, G. (2015). Research classification and the social sciences and humanities in Australia: (Mis)Matching organisational unit contribution and the impact of collaboration. *Research Evaluation*, *24*(3), 325-339. https://doi.org/10.1093/reseval/rvv006

Haddow, G., & Genoni, P. (2010). Citation analysis and peer ranking of Australian social science journals. *Scientometrics*, *85*(2), 471-487. https://doi.org/10.1007/s11192-010-0198-4

Haddow, G., & Hammarfelt, B. (2019). Early career researchers and evaluative metrics: Ambivalence, resistance and strategies. In F. Cannizzo & N. Osbaldiston (Eds.), *The social structures of global academia* (pp. 125-143). Routledge.

Haddow, G., & Noyons, E. (2013). Misfits: Research classification in research evaluation: Visualizing journal content within fields of research codes. In J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger, & H. Moed (Eds.), *14th international society of scientometrics and informetrics conference* (pp. 1210-1224). Australian Institute of Technology. http://hdl.handle.net/20.500.11937/7001

Haslam, N., & Koval, P. (2010). Possible research area bias in the excellence in research for Australia (ERA) draft journal rankings. *Australian Journal of Psychology*, *62*(2), 112-114. https://doi.org/10.1080/00049530903334489

Hicks, D. (2010). Overview of models of performance-based research funding systems. In *Performance-based Funding for public research in tertiary education institutions: Workshop proceedings* (pp. 23-52). OECD. http://dx.doi.org/10.1787/9789264094611-4-en

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden manifesto for research metrics. *Nature*, *520*(7548), 429-431. https://doi.org/10.1038/520429a

Ingwersen, P., & Larsen, B. (2014). Influence of a performance indicator on Danish research production and citation impact 2000-12. *Scientometrics, 101*(2), 1325-1344. https://doi.org/10.1007/s11192-014-1291-x

Khazragui, H., & Hudson, J. (2015). Measuring the benefits of university research: Impact and the REF in the UK. *Research Evaluation, 24*, 51-62. https://doi.org/10.1093/reseval/rvu028

Knott, M. (2015, November 17). Universities questioned over alleged 'gaming' of research rankings. *The Sydney Morning Herald*. https://www.smh.com.au/education/universities-questioned-over-alleged-gaming-of-research-rankings-20151117-gl0yva.html

Kwok, J. T. (2013). *Impact of ERA research assessment on university behaviour and their staff*. NTEU National Policy and Research Unit. https://apo.org.au/node/33769

Lamp, J. W., & Fisher, J. (2010). ERA distribution of information systems journals. *Australasian Journal of Information Systems, 16*(2), 5-17. https://doi.org/10.3127/ajis.v16i2.659

Larkins, F. (2019a, June 17). Do Australia's ERA discipline assessments really measure research excellence? *Times Higher Education*. https://www.timeshighereducation.com/blog/do-australias-era-discipline-assessments-really-measure-research-excellence

Larkins, F. (2019b). *Research at Australian universities: Is excellence really excellent?* The University of Melbourne. https://melbourne-

cshe.unimelb.edu.au/__data/assets/pdf_file/0003/3058221/2019_F-P-Larkins_ERA-Excellence.pdf

Larkins, F. (2019c). *Anomalies in the research excellence ERA performances of Australian universities*. The University of Melbourne. https://franklarkins.files.wordpress.com/2019/05/a35.-fp-larkins_university-performances-era-2018-12.pdf

Laudel, G., & Gläser, J. (2006). Tensions between evaluations and communication practices. *Journal of Higher Education Policy and Management*, *28*(3), 289-295. https://doi.org/10.1080/13600800600980130

Marsh, H., Smith, B., King, M., & Evans, T. (2012). A new era for research education in Australia? [Special issue]. *Australian Universities' Review*, *54*(1), 83-93. http://hdl.handle.net/10536/DRO/DU:30043049

Mingers, J., & Willmott, H. (2013). Taylorizing business school research: On the 'one best way' performative effects of journal ranking lists. *Human Relations*, *66*(8), 1051-1073. https://doi.org/10.1177/0018726712467048

Muhonen, R., Benneworth, P., & Olmos-Peñuela, J. (2020). From productive interactions to impact pathways: Understanding the key dimensions in developing SSH research societal impact. *Research Evaluation, 29*(1), 34-47. https://doi.org/10.1093/reseval/rvz003

Ossenblok, T. Lf. B., Engels, T. C. E., & Sivertsen, G. (2012). The representation of the social sciences and humanities in the Web of Science a comparison of publication patterns and incentive structures in Flanders and Norway (2005-9). *Research Evaluation, 21*(4), 280-290. https://doi.org/10.1093/reseval/rvs019

Pontille, D., & Torny, D. (2010). The controversial policies of journal ratings: Evaluating social sciences and humanities. *Research Evaluation*, *19*(5), 347-360. https://doi.org/10.3152/095820210X12809191250889

Research Excellence Framework. (2020). *Index of revisions to the 'guidance on submissions' (2019/01).* https://www.ref.ac.uk/media/1447/ref-2019_01-guidance-on-submissions.pdf

Ross, J. (2019, September 7). Fresh questions over Australian evaluation of humanities research. *Times Higher Education*. https://www.timeshighereducation.com/news/fresh-questions-over-australian-evaluation-humanities-research

Sheil, M. (2014). On the verge of a new era: Despite its limitations, excellence in research for Australia was the right assessment tool at the right time. *Nature, 511*(7510), S67. https://doi.org/10.1038/511S67a

Sivertsen, G. (2017). Unique, but still best practice? The Research excellence framework (REF) from an international perspective. *Palgrave Communications*, *3*(1), Article 17078. https://doi.org/10.1057/palcomms.2017.78

Vanclay, J. K. (2011). An evaluation of the Australian research council's journal ranking. *Journal of Informetrics*, *5*(2)*,* 265-274. https://doi.org/10.1016/j.joi.2010.12.001

Woelert, P., & McKenzie, L. (2018). Follow the money: How Australian universities replicate national performance-based funding mechanisms. *Research Evaluation*, *27*(3), 184-194. https://doi.org/10.1093/reseval/rvy018

Woelert, P., & Millar, V. (2013). The 'paradox of interdisciplinarity' in Australian research governance. *Higher Education, 66*, 755-767. https://doi.org/10.1007/s10734-013-9634-8

Woelert, P., & Yates, L. (2014). Too little and too much trust: Performance measurement in Australian higher education. *Critical Studies in Education*, *56*(2), 175-189. http://dx.doi.org/10.1080/17508487.2014.943776

Woodward, S. (2011, June 6). New journal system: ARC gets busy explaining. *Campus Review*. https://www.campusreview.com.au/2011/06/new-journal-system-arc-gets-busy-explaining/

Young, S., Peetz, D., & Marais, M. (2011). The impact of journal ranking fetishism on Australian policy-related research. *Australian Universities' Review*, *53*(2), 77-87.