Commissioners
Tertiary Education Quality and Standards Agency
standards@teqsa.gov.au.

Submission by Research Coaching Australia, Higher Education and Research Group and Research Strategies Australia:

TEQSA Consultation Process - Proposed Text for the Draft Legislative Instrument: Research

Dr T Cahill
Chief Executive Officer
Research Strategies Australia

Emeritus Professor K Houghton
Chief Academic Strategist
Higher Education and Research Group

M Clisby
Chief Executive Officer
Research Coaching Australia

28 April 2021.

Submission:

TEQSA Consultation Process - Proposed Text for the Draft Legislative Instrument: Research

Summary

The draft text for the proposed legislative instrument to make determination to an assessment of research and in particular quality is premised on a range of proxy measures for research quality and the validity, reliability and relevance of such measures. It is our view that the current wording will result in less than optimal outcomes, and that TEQSA would be well advised to limit its ambitions around 'measuring' research quality and expand its wording to consider research productivity. Measures of productivity can be filtered and weighted to allow a more nuanced and realistic research assessment approach.

Issue 1: The problems with bibliometric indicators

The problems with bibliometric indicators such as 'the volume of citations', and 'the quality of the publications in which those citations occurred' are well documented. This includes the following:

- While bibliometric indicators are widely used for research evaluation in STEM disciplines, they are not appropriate tools in HASS disciplines. The publishing behaviour in most STEM disciplines follows something like Bradford's Law – there is a core set of journals which constitute a discipline.[1] In HASS, the case is more complex as the core of each discipline is less apparent. In part, this is due to the prevalence of books (and book chapters), national and regional literature, and non-scholarly publications as primary modes of academic communication. This includes so-called non-traditional research outputs (NTRO), which includes research as varied as performance, museum catalogues, exhibitions, etc. In the latest round of ERA, for example, FoR 19 Studies in Creative Arts and Writing was comprised of only 40 per cent journal articles.[2] Aside from making the core difficult to identify, this poses an additional problem with respect to identifying and evaluating the relative value of different publication types. In both cases, HASS differs from the strictly hierarchical scientific publishing conventions in STEM which enable bibliometric analyses.
- There are significant issues of coverage of HASS literature in the major commercial databases of bibliometric data. Despite efforts over recent years to expand coverage of HASS journals by bibliometric databases such as Scopus and Web of Science (WoS), and to capture more diverse output types (e.g. books and book chapters), there is still a very small

[1] "journals in a single field can be divided into three parts, each containing the same number of articles: 1) a core of journals on the subject, relatively few in number, that produces approximately one-third of all the articles, 2) a second zone, containing the same number of articles as the first, but a greater number of journals, and 3) a third zone, containing the same number of articles as the second, but a still greater number of journals. The mathematical relationship of the number of journals in the core to the first zone is a constant n and to the second zone the relationship is n^2.Bradford expressed this relationship as 1:n:n." Potter, W.G. (1988). Of Making Many Books There is No End: Bibliometrics and Libraries. The Journal of Academic Librarianship, 14, 238a-238c.

[2] https://dataportal.arc.gov.au/ERA/NationalReport/2018/pages/section4/research-outputs-by-type/

proportion of the available literature represented in these databases. It is not much better in social sciences, where it has been estimated that less than 50 per cent of a researcher's portfolio appears in bibliometric databases.[3] Any research evaluation using citation data supplied by Scopus or WoS would be based on very small samples of an academic's work, and it is unlikely that this would provide for robust and verifiable assessment of their portfolio. The limited coverage means that commercial bibloimetric databases are not suitable for research evaluation in HASS.[4]

Issue 2: The problem of alternative data sources for bibliometric analyses

Related to the above are the challenges with the use of alternative databases (including such sources as Google Scholar and Microsoft Academic) which generally have a much higher, and arguably more valid coverage of HASS publishing. Regrettably there are challenges to this approach also apparent:

- Google Scholar (GS) is widely accepted as having a more comprehensive coverage of content in HASS, and a number of independent studies have concluded that the extent of indexing is high enough to suggest that GS may be a useful tool for bibliometric analysis in HASS. [5] Despite the coverage of GS – which is its distinct advantage over the other solutions – there is broad agreement that it is not mature enough to be used for bibliometric analysis currently. There are two primary reasons: first, there is a large amount of manual data processing that is required, and there is no way to access data and metadata; second, unlike Scopus and World of Science (WoS), which are closed systems where both cited-documents and citing-documents are known and limited, GS is an open system, and existing bibliometric methods are not necessarily suitable. Both Scopus and WoS are closed systems – that is, they only index a known number of journals (more or less the core journals of each discipline), and they only record the flow of citations between these journals. In this respect, bibliometric methods make comparisons amongst a group of more or less homogenous research papers, and can therefore distinguish between their relative performance given known parameters. Having a limited set of journals also allows for each journal to be pre-assigned to one or another discipline, which means that bibliometric methods can account for differences in citation practices between disciplines.

---

[3] See for example Amara & Landry 2012; Mingers & Lipitakis 2010.

[4] See for example Van Leeuwen, T. (2013). Bibliometric research evaluations, web of science and the social sciences and humanities: A problematic relationship? Bibliometrie-Praxis und Forschung, volume 2, pp. 8-18; and Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., et al. (2015). The metric tide: Report of the independent review of the role of metrics in research assessment and management. HEFCE, London.

[5] For example, Harzig (Harzing, A.W. (2014). A longitudinal study of Google Scholar coverage between 2012 and 2013. Scientometrics, volume 98, issue 1, pp. 565–575.) found that throughout 2012, GS expanded its coverage of certain disciplines, including economics, and that the rate of expansion has subsequently stabilised. Prins et al (Prins, A.A.M., Costas, R., van Leeuwen, T.N., & Wouters, P.F. (2016). Using Google Scholar in research evaluation of humanities and social science programs: A comparison with Web of Science data. Research Evaluation, volume 25, issue 3, pp. 264-270) similarly found that GS could be used as a tool for bibliometric analysis in Education, Pedagogical Sciences, and Anthropology. Meanwhile, Bornman (Bornmann, L., Thor, A., Marx, W., & Schier, H. (2016). The application of bibliometrics to research evaluation in the humanities and social sciences: An exploratory study using normalized Google Scholar data for the publications of a research institute. Journal of the Association for Information Science and Technology, volume 67, pp. 2778–2789) found that it might be more widely applicable as a tool across HASS disciplines by showing how results from GS converge with results from WoS. Finally, Martín-Martín et al ([5] Martín-Martín, A., Orduña-Malea, E., Ayllón, J. M., & López-Cózar, E. D. (2014). Does Google Scholar contain all highly cited documents (1950–2013)? arXiv preprint arXiv:1410.8464) have concluded that GS now indexes all the major sources of academic publication, including journal publishers, institutional repositories, digital hosts and scholarly societies. Harzing, A.W. (2013). A preliminary test of Google Scholar as a source for citation data: a longitudinal study of Nobel prize winners. Scientometrics, volume 94, issue 3, pp. 1057–1075.

This is a fundamental aspect of bibliometric analysis, given that different disciplines have different rates of citation. Scopus and WoS do not claim to capture every citation flowing to the articles in those journals, only those that are in the journal set they hold. This is very different to GS, which derives its citations from anywhere on the worldwide web, from an unknown number of sources, of unknown providence, and without any discipline classification.

- Microsoft Academic (MA), like GS, uses web search as the basis for indexing academic literature. As such, the advantage of MA is that it indexes more so-called non-source items – those that are not indexed in Scopus and WoS – such as book chapters, monographs and edited volumes. Compared with both Scopus and WoS, MA outperforms them significantly in this respect. It is important to note, though, that while these items are indexed more in MA than in the commercial databases, they are still at very low rates. For example, a recent study of MA coverage for outputs from the University of Zurich identified that, while Scopus and WoS indexed 6.2 per cent and 3.3 per cent of the university's monographs, respectively, MA indexed 15.1 per cent.[6] Although a vast improvement, this is still too small to be representative of the university's research. The same was true for edited volumes (15.6 per cent) and book chapters (14.3 per cent). The more important issue with MA as a data source for bibliometrics are the many problems that have been documented with its metadata. These include the vast number of 'fields of study' and issues with the accuracy of publication years.[7] In existing bibliometric analysis, citation performance relies on normalising for document type, discipline, and publication year. While there have been case studies where manual curation of data has been performed with MA, these have concluded that it is prohibitively time consuming, and cannot be scaled beyond the smallest cases. Without manual data curation, fundamental aspects of bibliometric analysis cannot be performed with MA.

Issue 3: Global push against journal-level metrics

In terms of the suggestion of journal level-indicators, globally there is a substantial push to move away from journal-level metrics such as journal impact factors as a proxy for the utility of individual research papers published therein. Specifically, the Declaration on Research Assessment (DORA) explicitly advises against this practice, and has been signed by approximately 2,200 organisations and greater than 17,000 individuals, including well respected funding councils including the National Health and Medical Research Council (NHMRC), publishers such as Springer Nature, citation data providers such as Elsevier (Scopus), individual universities including the University of Melbourne and individual academics. It is difficult to see how the proposed journal quality metrics proposed would be compatible with this global agenda.

[6] Hug, S.E. & Brändle, M.P. (2017). The coverage of Microsoft Academic: Analysing the publication output of a university. Retrieved from: https://arxiv.org/ftp/arxiv/papers/1703/1703.05539.pdf

[7] Citation analysis with Microsoft academic. Scientometrics (DOI: 10.1007/s11192-017-2247-8).

A Potential Solution: An alternative approach that brings greater objectivity and validity of coverage

There are valid and some would say compelling reasons to consider the use of Excellence in Research for Australia (ERA) results in whole or part. Funded and operated with a high level of integrity by a Federal Government agency, the ERA has overcome many of the problems outlined above around research quality assessments, and is widely accepted in Australia and abroad as best practice.

A serious issue with ERA is that it is only periodical, and that there are likely to be lags between registration of new universities and ERA assessment cycles. There are additionally questions associated with volume thresholds – ERA requires a minimum level of research output to be assessed under its current system.

We would therefore argue that as sensible option would be to have a productivity-based monitoring program in place on an annual cycle in combination with periodical ERA assessments to validate research quality where applicable. The more frequent productivity analysis can be informed by the periodic ERA outcomes. To that end, we submit that the Research and Education Efficiency Frontier Index (REEF-Index.com) developed by the Higher Education and Research Group (HERG) provides an innovative approach to analysis of research and teaching productivity in higher education.

The REEF Index is unique in the market. It measures the overall productivity performance of institutions against the two key 'outputs' of the university sector – Research in combination with Education. Importantly the REEF Index is neutral as to where an institution chooses to position itself in respect of the 'teaching – research' mix and assess research in an objective way. Importantly, it can employ either weights or filters on measures of research (as well as QUILT scores for teaching quality outcomes). That is to say, the ERA outcomes can be used to enhance and inform the research quality filters.

What we have in mind would be the creation of a broad quality measure framework, the calibration of the research expectations ideally tailored for the 4-digit Fields of Research (FoR), a proposed mechanism for keeping this framework and research expectations contemporaneous and an evaluation and feedback mechanism to ensure ongoing validity of the framework. Importantly, the advanced version of the REEF Index allows for individualized filtering and weighting of inputs. For example, research can be either filtered for certain levels of research quality, or can weight quality differences in ways that are valid for the specified task.

The REEF Index can be applied at multiple levels were relevant. It was originally used to track universities at the 'whole of institution' level. This also allows for comparisons between institutions. The second level of analysis was developed more recently and now permits research productivity analysis at a within institution level, including either academic unit (faculty, school or department) or at two or more usefully four digit FoR level. The third level of analysis is within the academic unit level – that is to examine outcomes at individual researcher level or for teams of researchers.

Depending on the data that a client might wish to employ, the REEF Index can 'harvest' research and scholarly outputs that are traditional research journal publications only or can use wider definitions of scholarly output activities. The data can also employ data around research grant income.

The advantage of this approach is that it is volume agnostic, can be run frequently, and allows for the relative researcher productivity to be objectively measured. This measurement can demonstrate

the research status of any university and can provide assurances that the category requirements are being met.

While the REEF Index has applications in a range of other jurisdictions, REEF was developed specifically for application in Australia. As such it allows for subtle nuances of the Australian system which is dominated by a relatively small number of largely comprehensive universities, with generally smaller and more specialised private providers.

References can be found to this analytic tool in the KPMG Report[8]. See also the other analyses provided in either the 2017 or 2018 *Higher Education Summits* and the specialist higher education press[9].

Concluding Remarks

Each of the three organisations involved in this submission seeks to support a robust, valid and reliable measurement of research. We have come together to provide this submission as we believe this may assist in the further development of a mature and strong higher education system for Australia.

Each of us is able to speak to this submission and are willing and able to speak to the contents of the submission. Should there be any questions or clarification, please feel free to contact us at any time.

[8] Parker, S., A. Dempster, M. Warburton, Reimagining tertiary education: From binary system to ecosystem KPMG, August 2018.

[9] AUSTRALIAN FINANCIAL REVIEW, August 27, 2018 'Universities make dramatic gains in efficiency led by the big research programs'
AUSTRALIAN FINANCIAL REVIEW, September 11, 2017 'Universities and Government dig in on funding cuts'
THE AUSTRALIAN, August 21, 2020 'Research productivity is growing; Teaching productivity is not'.
THE AUSTRALIAN, May 15, 2019 'On the efficiency frontier'.
THE AUSTRALIAN March 20, 2019 'Uncapping university efficiency'.
THE AUSTRALIAN, July 25, 2018 'How to optimise economies of scale in universities'.