



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Weatherall, KG;Giblin, R

Title:

Inoculating Law Schools Against Bad Metrics

Date:

2021-01-25

Citation:

Weatherall, K. G. & Giblin, R. (2021). Inoculating Law Schools Against Bad Metrics. Elsevier BV.

Persistent Link:

<https://hdl.handle.net/11343/326403>

# Inoculating law schools against bad metrics<sup>1</sup>

By Kimberlee Weatherall<sup>2</sup> and Rebecca Giblin<sup>3</sup>

For Professor Jill McKeough: who has inspired us not only to do rigorous research, but to make sure it reaches the people it matters to, and to work collaboratively, responsibly, and with joy.

## Introduction

Devastating consequences can flow from relying on proxies to measure university excellence. Take university rankings, for example. In the 1980s, Editors of the US News & World Report, a struggling magazine, thought creating a special college ranking edition might temporarily lift flaccid sales. Very quickly, these (rather arbitrarily-derived) rankings became critical to attracting enrolments and alumni donations, kicking off an arms race that, decades later, still has no end in sight. The magazine part of the business has long since been jettisoned, and the company instead churns out ratings of an ever-increasing range of organisations all around the world on an industrial scale. And university rankings have proliferated, as the US News & World Report has been followed by various other for-profit and private equity-owned competitors. They each rely upon various metrics to assess excellence, including factors like the percentage of applicants accepted for admission and employment rates after graduation. In the US in particular, colleges find themselves forced to pour resources into improving their performance on these proxies for excellence rather than improving students' educational experiences. To artificially drive down acceptance rates, they reject qualified candidates and accept excellent applicants they judge likely to accept an offer from a better school. To game other criteria (like the grade point average of enrolled students) some then expensively target a small subset of students using financial aid enticements that could otherwise have funded candidates with greater need. Universities routinely hire their own graduates briefly after graduation so they are 'employed' for ranking purposes, then drop them as soon as the crucial date has passed. Schools even wastefully devote resources marketing *to each other*, in an attempt to up their ranking's reputational component. One brand-new mathematics department in Saudi Arabia managed to get itself ranked ahead of heavyweights Cambridge and MIT by offering fat cash payments to highly-cited mathematicians who were willing to change their academic affiliations to include it, and staff at other universities have been caught fraudulently reporting false data as part of desperate attempts to keep up with their competitors, some of whom were probably cheating

---

<sup>1</sup> We'd like to acknowledge Francois Petitjean, Farrah Ahmed, Kate Seear, Matthew Harding, Katy Barnett, Gerard Goggin and Ksenia Sawczak for their generous and thoughtful feedback on an earlier draft, and Joshua Yuvaraj for research assistance.

<sup>2</sup> Professor of Law, The University of Sydney Law School. This work was supported in part by funding from the Australian Research Council (DP200101578).

<sup>3</sup> Associate Professor of Law and ARC Future Fellow, Melbourne Law School. This work was supported in part by funding from the Australian Research Council (FT170100011).

too. Factors that don't feed into the rankings machine (like degree cost, and long term student satisfaction with their educational experience) are neglected.<sup>4</sup>

Universities aren't the only ones that are rated and ranked. As academics, we are beset by systems demanding that we report our research achievements: when we apply for positions and promotions and/or tenure; in performance appraisals; when we apply for research grants or submit papers for journal publications or prestigious conferences. We're obliged to construct detailed reports at a faculty level as part of government processes for assessing universities such as the Australian Excellence in Research Assessment (ERA) or the UK's Research Excellence Framework (REF). In all of these contexts, academics are asked to report on their achievements, and, often, to devote labour towards assessing the achievements of others.<sup>5</sup>

Especially across humanities and social sciences (HASS) disciplines, much of this assessment continues to be based on peer review.<sup>6</sup> However, we are continually pushed towards greater use of metrics to enable researchers, departments and universities to be easily compared and rated. Meta-researchers have long critiqued this trend for moving us further towards a neoliberal audit culture and commoditisation of academic labour.<sup>7</sup> This has broad, debilitating effects on the research enterprise and academic freedom, adding to a culture of fear, precarity and overwork. As Bowrey puts it, 'metrics engender[] anxiety and insecurities so that academics continually scrutinise their own behaviour, choices and performance in terms of the norms and expectations of management.'<sup>8</sup> And efforts to 'do better' in terms of the metrics can end up diverting critical resources from areas of real need (as we showed in the opening paragraph), while doing little to further the reasons that

---

<sup>4</sup> Cathy O'Neill, *Weapons of Math Destruction* (Crown, 2016) 51-62.

<sup>5</sup> In Britain it has been estimated that 8% of university teaching budgets are spent on quality assurance: John Holmwood et al (eds), 'In Defence of Public Higher Education: Knowledge for a Successful Society' (Convention for Higher Education, May 2016) <<https://heconvention2.files.wordpress.com/2016/06/awp1.pdf>> 28. One *conservative* estimate of the 2018 ERA assessment was estimated to cost the government \$18 million, and the GO8 universities \$8 million in administrative costs - *not* counting opportunity costs and hours spent by academics in preparing returns, and as reviewers and assessors: Tim Dodd, 'Behind the numbers, an army of tweakers', *The Australian* (Canberra, 27 March 2019) 30. The ARC no longer collects data from universities on how long it takes universities to generate the necessary reports: Ksenia Sawczak, 'If it's all about transparency, show us ERA's hidden costs', *The Australian* (Canberra, 24 January 2018) 25.

<sup>6</sup> The Excellence in Research Assessment Exercise in Australia divides disciplines into two groups: those analysed by reference to *citation analysis* and those assessed by *peer review*. The disciplines are set out in the ARC's Discipline Matrix, but broadly, fields where at least half of the journals are indexed by the commercial citation analysis provider (Clarivate in ERA 2018) use citation analysis. These disciplines are primarily STEM disciplines. Humanities and Social Sciences (including law) are assessed via peer review: that is, the university nominates 30% of its outputs in each category (books, journal articles etc) and these are sent to peer assessors nominated by other universities. These ratings are used alongside other indicators and explanatory statements to assign a rank (1-5) to the research discipline at each university (if it meets the volume threshold): see Australian Research Council, *ERA 2018 Evaluation Handbook* (2018), Chapter 4.

<sup>7</sup> See eg Richard Watermeyer and Adam Hedgecoe, 'Selling 'impact': peer reviewer projections of what is needed and what counts in REF impact case studies. A retrospective analysis' (2016) 31(5) *Journal of Education Policy* 651.

<sup>8</sup> Kathy Bowrey, 'Audit Culture: Why Law Journals Are Ranked and What Impact This Has On The Discipline of Law Today' (2013) 23(2) *Legal Education Review* 291, 294.

universities exist: to advance and disseminate knowledge, and by doing so help address society's many critical challenges.

Meta-researchers and scientometricians (those who understand scholarly metrics best) warn they're increasingly being misused. The 2012 San Francisco Declaration on Research Assessment (DORA) calls for 'the *responsible* use of metrics that align with core academic values and promote consistency and transparency in decision-making' (emphasis added).<sup>9</sup> The 2015 Leiden Manifesto was developed in response to the trend of research evaluations being led by data, rather than judgment, and concerns that 'We risk damaging the system with the very tools designed to improve it, as evaluation is increasingly implemented by organizations without knowledge of, or advice on, good practice and interpretation.'<sup>10</sup> These high level principles and declarations are useful and important, but if we want them to be reflected in our sector, we need ways to implement them in practice.<sup>11</sup>

In the discipline of law and legal studies in Australia, we have seen the dangers of inappropriate reliance on metrics up close, most notably via the disastrous 2010 ERA journal list. Expert law academics had universally agreed that journal rankings weren't suitable for assessing legal research, given the discipline's size, idiosyncratic journal publishing structures,<sup>12</sup> high degree of specialisation, and the reality that, as a similar exercise in the UK had established, legal research of the highest quality can be found in a wide range of journals.<sup>13</sup> Despite law's courteous and steadfast resistance, however, when the Australian Research Council (ARC) released its draft journal ranking in mid 2008, law journals were included. The list was almost ludicrously inappropriate, featuring just two non-US law journals in the top 198 (despite the fact that few US journals would even consider publishing research on Australian law), and with the ARC apparently unaware that most of them were not subject to peer review. But still, it had to be taken seriously. As Bowrey explains, 'from this time on there was no longer any real discussion of rejecting rankings but rather only discussion of how to 'improve' a bad situation.'<sup>14</sup> The eventual list of law journals was amended, but it remained highly unsatisfactory. While some inequities were fixed during the revision process, fresh ones were introduced thanks to the fierce lobbying and rent-seeking that took place once it became clear that rankings would be used to assess legal research and law schools.<sup>15</sup>

---

<sup>9</sup> *San Francisco Declaration of Research Assessment* (Web Page) <<https://sfdora.org/>>.

<sup>10</sup> Diana Hicks et al, 'Bibliometrics: The Leiden Manifesto for research metrics' (2015) 520 (7548) *Nature* 429, 429.

<sup>11</sup> Moher et al note that most proposals (especially those noting the importance of assessing broader contributions like collaboration, mentoring etc) "were silent on details of how exactly faculty assessment committees could implement their proposals, what barriers to implementation existed, and how to overcome them": David Moher et al, 'Assessing scientists for hiring, promotion, and tenure' (2018) 16(3) *PLoS Biology* e2004089, 10.

<sup>12</sup> In particular, many US law journals are university (rather than subject-matter) based, student-edited and not peer-reviewed, with the prestige of the journal determined largely by the prestige of the university which publishes it.

<sup>13</sup> RAE 2001, *RAE 2001 - Overview reports from the panels* (Web Page) <<http://www.rae.ac.uk/2001/overview/>>, Report 36 (Law).

<sup>14</sup> Kathy Bowrey, 'Audit Culture: Why Law Journals Are Ranked and What Impact This Has On The Discipline of Law Today' (2013) 23(2) *Legal Education Review* 291, 302.

<sup>15</sup> See Ian Murray and Natalie Skead, "'Who Publishes Where?': Who publishes in Australia's top Law journals and which Australians publish in top Global Law journals?' (2020) 47(2) *University of Western Australia Law Review* 220, 226 - 227 for a fuller discussion of criticisms of the 2010 list.

While publishing in top ranked journals was supposed to be a reasonable proxy for publishing the best research, it ended up rewarding all kinds of other characteristics instead. The highest ranked Australian law journals were ultimately mostly generalist university reviews, which disproportionately publish men over women and public law over any other specialty<sup>16</sup> (indeed, one of the most prestigious journals, the *Federal Law Review*, has announced that it will *only* publish public law (federal or state), abandoning a long-standing position of publishing the best work on areas within the federal jurisdiction). Research that was interdisciplinary, international, or covered one of the (many) snubbed sub disciplines could be incapable of finding a home in a top ranked journal, even if it was important, urgent, and of excellent quality. The reality is that editors of the leading law school journals often feel unqualified to review or publish interdisciplinary work in particular, or work from smaller sub-disciplines they judge ‘not of general interest’. And the use of the flawed list wasn’t limited to the research assessment exercise: as had been predicted, universities rapidly co-opted those rankings to inform all manner of other processes, including promotions, recruitment and internal grant allocations. Part of the list’s distorting effect was to change where researchers submitted their research - and even the topics on which they worked. For example, one empirical study of industrial relations academics found they were changing the focus of their work in order to bring it within the purview of those higher ranked journals by making it less Australia-focused and less critical.<sup>17</sup>

Such distortions shouldn’t have been surprising: when Australian research was funded using a formula based on the number of papers published, it caused the number of papers to go up, but they were in less-cited journals - suggesting that researchers increased their rate of productivity at the cost of quality.<sup>18</sup> But some people *were* surprised. The ARC’s Chief Executive noted with alarm that, while the rankings were only ever intended to be a minor part of the assessment exercise, ‘universities are using it in ways that are more rigid than I would [have] ... [and] other than what was intended.’<sup>19</sup> With their flaws increasingly obvious, the journal rankings were only ever used in one ERA exercise before being abandoned. However, that hasn’t been the end of their influence. The 2010 journal list is *still today* widely used within Australian law schools to assess research quality, and by researchers in grant applications seeking to demonstrate their publications’ quality,<sup>20</sup> despite the ARC telling people not to. That’s the case even though, having been dumped a decade ago, the list’s panoply of problems now includes ‘hugely out of date’.

---

<sup>16</sup> Katy Barnett, ‘Citation as a measure of impact: Female legal academics at a disadvantage?’ (2019) 44(4) *Alternative Law Journal* 267, 269-271; Katy Barnett, ‘High Court citation as a measure of ‘impact’: questions and challenges’ (unpublished conference paper, 2020, 12).

<sup>17</sup> Suzanne Young, David Peetz, Magalie Marais, ‘The impact of journal ranking fetishism on Australian policy-related research: A case study’ (2011) 53(2) *Australian Universities’ Review* 77, at 80-83.

<sup>18</sup> Linda Butler, ‘Explaining Australia’s increased share of ISI publications - the effects of a funding formula based on publication counts’ (2003) 32(1) *Research Policy* 143.

<sup>19</sup> (Annabel McGilvray, ‘Nervousness over research ratings’ *Campus Review* (Blog Post, 27 September 2010) <<https://www.campusreview.com.au/2010/09/nervousness-over-research-ratings/>>.

<sup>20</sup> Ian Murray and Natalie Skead, ‘“Who Publishes Where?”: Who publishes in Australia’s top Law journals and which Australians publish in top Global Law journals?’ (2020) 47(2) *University of Western Australia Law Review* 220, 226-227.

This history shows us that, when a form of auditing is mandated with sufficient force, and there's no coherent, well-accepted alternative in place, it can be almost impossible to resist. We can think of it as being a bit like the human microbiome. A healthy gut is filled with beneficial bacteria, which leaves little room for bad ones to get established. In an unhealthy or underpopulated gut however, undesirable bacteria can quickly take root and take over. We don't mean to imply that metrics take over and spread all on their own (that would let our profession off the hook!) - human agency certainly plays a key role in giving them value and meaning. Rather, our point is that, once the research ecosystem is populated with bad metrics, it can be very difficult to oust them in favour of something better.

The lingering influence of journal rankings in law exemplifies the point. Before the ARC released its draft, there was near universal agreement within law schools that journal ranking was unsuitable for the discipline, for all the reasons we canvassed above. However, once it was introduced, it took hold. Now, the discipline's focus has shifted towards improving the rankings, rather than doing away with them altogether. Despite resistance from some law schools, the creation of a revised and updated, 'better' rankings list has been official policy of the Council of Australian Law Deans since 2018, even though the original objections remain as valid as ever, and indeed, have new support from scientometricians concerned about metric misuse, as well as initiatives like DORA. Consultations undertaken in order to develop a new journal list triggered similar lobbying to that which we saw in the leadup to the 2010 list, with individual academics feeling forced to set aside their objections and participate to prevent their subdisciplines from being disadvantaged. At time of writing, we understand that the project to create a new list is ongoing.

In this contribution, we want to promote a conversation about what it might look like for academic researchers working in law faculties or on legal issues to assess research contributions that promote the shared values of the legal academy. Our focus is on two areas of research assessment: research impact and the bucket of concepts variously described as mentorship/supervision/leadership. There's a certain pragmatism about these foci. Although as law academics we are continually being asked to audit ourselves in different ways, the metric tide has not yet overwhelmed these two areas. But there are some signs of a tendency towards quantification already in evidence. And we suspect that both areas *could* be reduced to a capitalist-style production value by university management determined to extract, and demonstrate, ever-increasing productivity.

So we think it is critical both to push back against such quantification, and to articulate a more nuanced perspective. We argue that, by taking the initiative to do so as a discipline now, researchers will be empowered to create and interpret those narratives broadly, and with a focus on quality rather than metrics. In making these suggestions, we emphatically are *not* suggesting a 'one size fits all approach', such as 'be well rated by your peers', 'publish in these journals', 'get cited', or 'have impact'. To adapt the words of Paul Feyerabend into the research assessment context: 'The only principle that does not inhibit progress is: anything goes.'<sup>21</sup> However, we want to draw attention to the fact that the power to decide what is important to us is in *our* hands: at least for now. Since we, as academics, provide almost all the assessment labour that gives meaning to these new measures of

---

<sup>21</sup> Paul Feyerabend, *Against Method: Outline of an Anarchistic Theory of Knowledge* (Humanities Press, 1975) 14.

performance, we have some power to decide how it's to be done, and to articulate and assess what we consider valuable. And, to the extent any attempt is made to colonise the discipline via a top-down metrics-based approach in these areas, it will have less room to take root. We can and should use this opportunity to advance the broader project of moving towards what Raewyn Connell calls the Good University; one that is creative, inclusive and democratic, and which fights against 're-masculisation' of universities (which partly manifests in a managerial culture 'that is aggressive, competitive, self-centred and emotionally cold'<sup>22</sup>).

## Research impact

### Impact or translation?

The development of new forms of research assessment has been a bit like the children's book, 'The King, the Mice and the Cheese.' That story begins with a mouse problem, which causes the King to bring in cats, which fixes the mice all right, but creates a cat problem, so he has to bring in dogs. And so it goes. Assessing research via measures of peer review and academic reputation led to reliability concerns. Bibliometrics (reporting on 'more objective' indicators like citation counts, h-indexes and impact factors), were introduced to address them, but were quickly themselves recognised to be narrow and damaging. That triggered the creation of alt-metrics - a broader set of metrics, which come with their own biases and problems - as well as the idea of evaluating researchers on their work's impact.

Assessment of impact was part of the initial design of research assessment in Australia,<sup>23</sup> and over time has taken hold as a core evaluation consideration in Britain (via the Research Excellence Framework, or REF) and in Australia, at both individual and institutional levels. In Australia, the Engagement and Impact Assessment 2018 (EI2018) was a first attempt to formalise such assessment at the university level. Like the ERA, the EI2018 assessed individual subject areas within universities.<sup>24</sup> Each research unit submitted three separate reports. Two parts sought to assess research impact: (1) the impact case study: a narrative<sup>25</sup> which might, at the option of the university, include quantitative metrics of impact,<sup>26</sup> and (2) a narrative describing the unit's approach to (promoting) impact (narrative). Part 3 sought to

---

<sup>22</sup> Raewyn Connell, *The Good University: What Universities Actually Do and Why It's Time for Radical Change* (Bloomsbury Publishing, 2019) 128.

<sup>23</sup> Ksenia Sawczak, 'Assessing Impact Assessment – What Can Be Learnt from Australia's Engagement and Impact Assessment?' *Impact of Social Sciences* (blog), May 16, 2019. <https://blogs.lse.ac.uk/impactofsocialsciences/2019/05/16/assessing-impact-assessment-what-can-be-learnt-from-australias-engagement-and-impact-assessment/>; Gerard Goggin, 'Research Engagement and Digital Humanities Interfaces in Open Communication Research', unpublished manuscript (copy on file with Weatherall).

<sup>24</sup> A subject area was defined as a 2-digit area as classified under the Australian Bureau of Statistics. 1297.0: *Australian and New Zealand Standard Research Classification (ANZSRC) - Fields of Research*, 2008. New FOR codes were issued in 2020. In the 2008 Classification, Law/Legal studies was its own 2-digit FOR (1801), one of a total of 22 2-digit FOR codes. Some 2-digit codes were much broader: for example 1601 ('Studies in Human Society'), which includes anthropology, political science, sociology, criminology, and others. Universities also had the opportunity to submit impact case studies specifically focused on interdisciplinary research and research for or about indigenous people or issues.

<sup>25</sup> Australian Research Council, *EI 2018 Assessment Handbook* (2018) 8.

<sup>26</sup> Australian Research Council, *EI Submission Guidelines* (2018), Australian Research Council, *EI 2018 Assessment Handbook* (2018) 35-36.

assess broader societal engagement with research, using quantitative data the ARC already held about research funding received from the private and public sector or foundations, and a narrative describing other evidence of engagement, in which units could also propose their own quantified evidence of engagement.

EI2018 has been criticised for being too limited, too subjective, and insufficiently rigorous,<sup>27</sup> as well as being too susceptible to the spin and rhetoric of a beautifully written narrative. Now the king has an impact problem. And, as with the story about journals we described above, one answer proposed to the rigour and subjectivity problem is more metrics.

The limited data used EI2018 - only in relation to engagement - were clearly unsatisfactory, although it's easy to see why the ARC decided to feed data about external research funding into the analysis: it is easily digestible, and even more importantly, the data were already on hand. It was, in other words, a 'metric of convenience' and preferred by universities already fatigued with the constant reporting of data, and unsure whether they *had* other reliable data stretching back 6 or more years. But using funding as a proxy for engagement likely disadvantages the kinds of law schools with the strongest focus on social justice research (who are also the likeliest to be working with cash-strapped NGOs and public sector organisations rather than corporates). It also fails to capture the extensive contributions legal scholars make to government and policymaking via law reform processes and informal advice. Conversely, it advantages those who have the best business development, alumni networks, and support for contracts and accounting. This is where university 'privilege machines' really kick into high gear.

Other metrics have been suggested for establishing engagement with research, although they all have their challenges and weaknesses. For example, another suggested metric of engagement was the number or proportion of papers co-authored with non-academics/industry. But this carries risks, for lawyers, of giving rise to perceived (or actual) conflicts of interest. Independence is important for all researchers, but we need to be particularly wary of encouraging legal researchers to do things that could cause them to be viewed as shills for particular industries or market actors. The Australian Academy of the Humanities has suggested that indicators directed to exhibitions and performances, and 'official engagement in government policy instrumentalities', although such metrics carry their own questions. If we transform these phenomena into numbers, are we now looking to count 'bums on seats'? Are exhibitions or seminars with smaller attendances less valuable? And are we all going to be even more overwhelmed by invitations to events and talks than we already are?

Assessing engagement with research lends itself more to numbers and common metrics than impact for now, and even those calling for more rigour in the assessment of impact have acknowledged that there are no universal metrics for doing so.<sup>28</sup> In Australia, the

---

<sup>27</sup> See, eg, Queensland University of Technology, *Submission in Response to the ERA EI Review Consultation Paper*, October 2020, Available at [https://cms.qut.edu.au/\\_data/assets/pdf\\_file/0016/1002292/19-October-2020-QUT-Response-to-the-ARC-ERA-EI-Review-2020\\_6584.pdf](https://cms.qut.edu.au/_data/assets/pdf_file/0016/1002292/19-October-2020-QUT-Response-to-the-ARC-ERA-EI-Review-2020_6584.pdf).

<sup>28</sup> Queensland University of Technology, 'Submission in Response to the ERA EI Review Consultation Paper', October 2020. [https://cms.qut.edu.au/\\_data/assets/pdf\\_file/0016/1002292/19-October-2020-QUT-Response-to-the-ARC-ERA-EI-Review-2020\\_6584.pdf](https://cms.qut.edu.au/_data/assets/pdf_file/0016/1002292/19-October-2020-QUT-Response-to-the-ARC-ERA-EI-Review-2020_6584.pdf).

narrative form currently dominates impact assessment, especially in the HASS disciplines. But across the board, we are seeing pressure to quantify impact. At the individual level, nationally competitive grants funded by the ARC and NHMRC require applicants to describe their “research achievements and contributions to the field” without mandating any specific metrics. In these sections, the applicant is encouraged to articulate in narrative form the beneficial impacts of their work outside of academia and other professional activities. These fields typically leave significant space for nuance and inclusive interpretation, enabling applicants to articulate in their own way the benefit their research has contributed to the economy, society, the environment, and culture. Still, in our own assessments of these grants, we have noted an increased tendency for candidates to rely on metrics, such as citation counts.

In EI2018, too, the writers of case studies sought to quantify their impact. Law emphasised its contributions to judicial training and processes through efforts like writing bench books – the ‘manuals’ that judges use to inform them of law, practice, and the latest research. Critically, they *quantified* this - XX judges used the bench books, resulting in XX efficiencies across the court system with XX improvements for litigants. Or they talked about how their research-led and informed clinics provided direct access to justice for disadvantaged communities - quantifying the number of vulnerable communities and people assisted.<sup>29</sup> Since Australia and the UK have a history of adopting each others’ ideas in research assessment, any further elaboration of metrics in the UK should sound early warning bells.<sup>30</sup> As research assessments grow more common and accepted, we fear that, as a discipline, we risk, over time, fetishizing proxies for impact in much the same way we began to fetishize journal rankings in 2010 (and still do so today). Indeed, the enduring influence of the 2010 ranking list may have *normalised* the use of proxies to assess legal research.

While proxies often look neutral on their face, they can actually entrench bias and patterns of discrimination - exactly as they did with the journal ranking list. The literature, for example, suggests that newspaper and media mentions of research may suffer gender bias.<sup>31</sup> Or more specifically for law, consider judicial citations, for example. In her analysis of three recent years of High Court decisions, Barnett found Australia’s High Court disproportionately cites men over women (505 to 49), mostly cites textbooks and treatises, and is more likely to

---

<sup>29</sup> This tendency has also been seen formally in the UK, where the Higher Education Funding Council for England (HEFCE), on behalf of the four UK higher education funding bodies, commissioned work to (i) identify quantitative indicators used as evidence of impact in the case studies submitted to the Research Excellence Framework (REF) 2014; and (ii) develop guidance for how these indicators could be standardised for potential use in REF 2021. That suggests a continued goal of coming up with measures that enable broad comparison and analysis: Sarah Parks et al, ‘Guidance for standardising quantitative indicators of impact within REF case studies’ (RAND Europe, 2018).

<sup>30</sup> References to metrics in UK REF2021 Submission Guidelines are muted; the submission refers to the primacy of expert assessment and peer review. It also provides an Equality Briefing to inform panels regarding the potential gender impacts of various methods of assessment: Research Excellence Framework: Equality Briefing for Panels, July 2018. [https://www.ref.ac.uk/media/1017/equality-briefing-for-panels-ref-2018\\_05.pdf](https://www.ref.ac.uk/media/1017/equality-briefing-for-panels-ref-2018_05.pdf). But this raises a question about what happens if Australia were to adopt UK metrics *without* the rest of the guidance that sits around it.

<sup>31</sup> Amberg, Amanda, and Darren N. Saunders. “Cancer in the News: Bias and Quality in Media Reporting of Cancer Research.” *PLOS ONE* 15, no. 11 (November 9, 2020): e0242133. <https://doi.org/10.1371/journal.pone.0242133>.

cite academics in public law than in any other field.<sup>32</sup> Holding up judicial citation as an archetypal or valued form of impact means prioritising certain kinds of research (particularly less risky doctrinal research) and certain kinds of researchers (men; those from the most prestigious law schools) over others. More theoretical and experimental research, including feminist and critical race theory, would suffer under such a system, as would genuine diversity of ideas and in the academy. Casting further doubt on the use of judicial citations as a proxy for research impact, a US analysis has found that courts mostly refer to descriptive elements of academic writing, and that the author's reasoning is only rarely influential.<sup>33</sup> Given what we know about how incentives can change researcher behaviour, all this suggests that overweighting judicial citations won't encourage much of the research for which we think the legal academy exists.

Citations are another potential metric for impact: scholarly citations as evidence of research quality, or citations in grey and policy literature as evidence of broader non-academic impact. We are unaware of any research on biases in citation in the grey or policy literature, but there is a large body of research on scholarly citation practices that should give us serious pause. Harrison and Mashburn found legal academics cite other academics on descriptive points to an even greater extent than their judicial colleagues: something that we should certainly take into account when considering what weight we want to give to citation numbers as a measure of 'impact' in and of themselves. They further found citations to be highly correlated with the prestige of the journal in which the paper was published, which, if it holds equally true in Australia, imports all the biases discussed in our introduction to boot.<sup>34</sup> Rewarding the number of citations also favours researchers with larger networks: often those with greater capacity to travel and attend evening events, which, given the distribution of caring responsibilities in society, are again more likely to be men and older academics. It also benefits those working in fields with the most research (or policy) activity, discouraging work on emerging, niche or local issues that might nonetheless be deserving of academic interrogation. On top of all that, men have been shown to cite men more than women, even in fields and journals which mostly are mostly populated by women.<sup>35</sup> Such (likely

---

<sup>32</sup> See Katy Barnett, 'High Court citation as a measure of 'impact': questions and challenges' (unpublished conference paper, 2020). This is in a context where a women constitute a slight *majority* of all legal researchers in Australia (662 compared to 601 men): Australian Research Council, *State of Australian University Research 2018/2019: ERA National Report* (Web Page) ['Gender'] <<https://dataportal.arc.gov.au/ERA/NationalReport/2018/pages/section1/gender/>>.

<sup>33</sup> Jeffrey Harrison and Amy Mashburn, 'Citations, Justifications, and the Troubled State of Legal Scholarship: An Empirical Study' (2015) 3 *Texas A&M Law Review* 45, 71.

<sup>34</sup> It should perhaps be noted that the paper studied citations in US law journals. This is relevant because US law journals are student-edited, and notorious for student editorial teams adding citations for even obvious points, at times without consultation with original academic authors. We can also speculate that student editors who are less familiar with the overall literature and leading thinkers may be unduly influenced by the prestige of the publication in choosing what to cite.

<sup>35</sup> See eg Michelle L Dion, Jane Lawrence Sumner, Sara McLaughlin Mitchell, 'Gendered Citation Patterns across Political Science and Social Science Methodology Fields' (2018) 26(3) *Political Analysis* 312. See also Daniel Maliniak, Ryan Powers and Barbara F. Walter 'The Gender Citation Gap in International Relations. International Organization, Available on CJO 2013 doi:10.1017/S0020818313000209; Silvia Knobloch-Westerwick and Carroll J Glynn, 'The Matilda Effect—Role Congruity Effects on Scholarly Communication' (2013) 40(1) *Communication Research* 3; Knobloch-Westerwick S., Glynn C.J., and Huges M., 'The Matilda Effect in Science Communication: An Experiment on Gender Bias in Publication Quality Perceptions and Collaboration Interest. *Science Communication*, 2013. 35(5): p. 603–625; Rossiter, Margaret W.

unconscious) biases drive down citations for some researchers. Book reviews can demonstrate meaningful engagement with a work, but in law, these are often commissioned from friendly scholars by authors or publishers, which means that engagement might not be worth as much as it might seem on its face. And again, reviews advantage the more populated research fields, more prestigious publications, and more established researchers.

Newer ‘altmetrics’ such as social media engagement will also disproportionately benefit men since attacks on women, people of colour, and members of LGBTQI+ communities have become commonplace.<sup>36</sup> They also disadvantage older researchers who are less likely to be comfortable using those tools for engagement, and researchers in subdisciplines without active social media communities. In short, *none* of these proxies can satisfactorily demonstrate impact in and of themselves, and most carry risks of disadvantaging the same groups over and over again, with compounding effect.

The use of proxies isn’t the only problem with current approaches to impact assessment. As Greenhalgh et al have found, in fields involving public policy making, ‘the links between research and impact are complex, indirect and hard to attribute’.<sup>37</sup> Although it’s notoriously difficult to link research to impact in the public policy sphere, some research assessments more or less demand that we do so. The ARC Future Fellowship rules in 2016 (as one example) required candidates to include a page-long statement of research impact and contributions to the field, describing ‘how [their] research has led to a significant change or advancement of knowledge’. This invites candidates to individualise and overstate their impact claims, disadvantaging those who are not willing to do so (as well as those who don’t have access to the mentoring to teach them how to find the *appropriate* level of individualisation and overstatement!). It would not be surprising to find that people of certain genders and cultural backgrounds are more comfortable claiming group achievements as their own than others.

Even where the instructions are not so explicit, the whole structure of impact and engagement assessment in Australia tends to push towards oversimplifying and overclaiming. In the EI2018, research impact was defined by the ARC broadly as ‘the contribution that research makes to the economy, society, environment or culture, beyond the contribution to academic research’.<sup>38</sup> In theory, this language is broad enough that quantifiable change in the world need not be established. Nevertheless, the Australian distinction between engagement and impact tends to mean that evidence of contributions to a broader public debate can tend to be characterised as engagement, and that measurable change or activity based on the research is treated as impact. Watermeyer and Hedgecoe discovered something similar in the context of the UK’s REF, observing that ‘the certainty of

---

“The Matthew Matilda Effect in Science.” *Social Studies of Science* 23, no. 2 (May 1, 1993): 325–41. <https://doi.org/10.1177/030631293023002004>;

<sup>36</sup> Regarding the ways such attacks can disproportionately affect female academics, see Emma Kavanagh and Lorraine Brown, ‘Towards a research agenda for examining online gender-based violence against women academics’ (2019) *Journal of Further and Higher Education*; Heather Savigny, ‘The Violence of Impact: Unpacking Relations Between Gender, Media and Politics’ (2019) 18(2) *Political Studies Review* 277.

<sup>37</sup> Trisha Greenhalgh et al, ‘Research impact: a narrative review’ (2016) 14(78) *BMC Medicine*, 2.

<sup>38</sup> Australian Research Council, *Engagement and Impact Assessment 2018-19: National Report* (Web Page) [‘Introduction: Definitions’] <<https://dataportal.arc.gov.au/EI/NationalReport/2018/>>.

impact attribution hinged on authors' proclivity in decollectivising, individualising and/or colonising impact claims or in other words, securing the exclusivity of bragging rights. Therefore, only with the magnification of the author's impact contribution, ostensibly with the nullifying of the (now evanescent) contribution of collaborators, could the claims being promulgated be treated confidently as authoritative declarations.<sup>39</sup> It is challenging to ensure narratives are not dominated by case studies that can adopt a linear logic: step 1 - conduct research, step 2 - take it to government, step 3 - look, impact!<sup>40</sup> Reichard et al found evidence in the context of the UK's equivalent 2014 REF that "High-scoring case studies appear to have conformed to a distinctive new genre of writing, which was clear and direct, and often simplified in its representation of causality between research and impact, and less likely to contain expressions of uncertainty than typically associated with academic writing".<sup>41</sup>

This push is unfortunate, because like all the measures that came before it, measurable societal, economic, or cultural change is effectively a proxy. Change is often not within the control of, or entirely or even mostly attributable to, any individual researcher or group,<sup>42</sup> and really important research may play only a small part in a long trajectory of improvement. Today thousands of climate researchers are researching, publishing, and engaging, and the work is having critical but insufficiently rapid impact. But no one individual or group could claim credit for the policy change that has been achieved, nor could they be blamed for humanity's halting progress. As Greenhalgh et al explain, 'A nuanced narrative may be essential to depict the non-linear links between upstream research and distal outcomes and/or help explain why research findings were not taken up and implemented despite investment in knowledge translation efforts.'<sup>43</sup> The challenges in providing such nuance is compounded by space constraints which limit researchers' ability to tell nuanced stories. EI2018 limited researchers' descriptions of their impact to 6000 characters (about 700 words, or one page), plus an 800 character summary and a list of just 10 publications.<sup>44</sup> ARC application forms have sometimes asked grant applicants to explain the impact or significance of publications in only 30 words. This false economy encourages metrics and outlaws nuance. Skimping means incentivising the wrong things. If we are to assess impact well, that means with nuance, and that requires space to do so.

Demanding that researchers show actual impact in the form of provable, attributable change (which is often largely out of their control) also encourages a focus on 'short term, proximal

---

<sup>39</sup> Richard Watermeyer and Adam Hedgecoe, 'Selling 'impact': peer reviewer projections of what is needed and what counts in REF impact case studies. A retrospective analysis' (2016) 31(5) *Journal of Education Policy* 651, 655.

<sup>40</sup> Gemma Derrick, *The Evaluators' Eye: Impact Assessment and Academic Peer Review* (Palgrave Macmillan, 2018) [Chapter 4].

<sup>41</sup> Bella Reichard et al, 'Writing impact case studies: a comparative study of high-scoring and low-scoring case studies from REF2014' (2020) 6(31) *Palgrave Communications* 1, 15.

<sup>42</sup> The situation might be somewhat different at the university or discipline level: a university researching across 22 broad disciplinary groupings that cannot point to a single real world change resulting from its research might have questions to answer. But EI2018 assessed at the level of the 2-digit FOR, not the university level overall.

<sup>43</sup> Trisha Greenhalgh et al, 'Research impact: a narrative review' (2016) 14(78) *BMC Medicine*, 4.

<sup>44</sup> Australian Research Council, *EI Submission Guidelines* (2018) 46. The UK's guidelines for REF2019/01 are similarly limited: 750 words for the details of the impact: REF2021, *Index of revisions to the 'Guidance on submissions' (2019/01)* (October 2020) 97 <[https://www.ref.ac.uk/media/1447/ref-2019\\_01-guidance-on-submissions.pdf](https://www.ref.ac.uk/media/1447/ref-2019_01-guidance-on-submissions.pdf)>.

impacts’ which ‘could create a perverse incentive against more complex and/or politically sensitive research whose impacts are likely to be indirect and hard to measure.’<sup>45</sup> Thus, it discourages work that is hard to translate because of the complexity or difficulty of the problem, including rigorous empirical or deep historical work and work on interdisciplinary, interconnected problems. McKeough’s work heading up the Australian Law Reform Commission enquiry into copyright and the digital economy is one such example. McKeough led the enquiry, spending two years painstakingly gathering and analysing evidence in this complex and politically charged area before releasing the final report in 2014. While some of the impact is obvious and linear - the work of McKeough and her team was, for example, directly proximate to the adoption of new fair dealing exceptions for libraries and archives and for people with disabilities - much of it is not. For example, the report brought consensus that reform was necessary (if not consensus about the shape it should take!), which shifted the copyright discourse forward. That’s difficult to attribute and prove, but a vital contribution nonetheless. Other important impacts are even more amorphous. For example, the value and breadth of Professor Jill McKeough’s contributions to society, research and research culture have provided inspiration and a model for our own careers, and influences the way we engage with ECRs and train our students. Jill will probably never again be audited for research impact, but if she is, she can cite this paper as proof of those intergenerational effects.

The ALRC inquiry also highlights the dangers of only over-rewarding work where the impact can be shown within reporting timelines, which can be narrow: 6 years, for example, in EI2018, based on associated research over 15 years.<sup>46</sup> The pace of law reform can be glacial. The McKeough report has been slowly and incrementally influencing copyright policy for seven years, and its impact looks set to continue for perhaps decades more. Any impact assessment framework that mandates short timelines risks disincentivising such vital work. And of course, junior researchers are of course most disadvantaged by the long pathways to impact as they are being repeatedly assessed before their work has had time to percolate, before their reputations are established, and before their networks are formed.

## Asking the right question: what have you done about your discovery?

We believe that research translation *is* critical, and an integral part of an academics’ job. We believe strongly that all researchers ought to be able to articulate why they are doing the work they do, why it matters, and how they are making the connection to ensure that the research they generate does reach those to whom it is (or should be) important. Academia has increasingly limited resources and the world is facing ever more urgent problems. But researchers themselves are best placed to decide what kind of translation best suits their work, and when. That is, translation should be seen as a dimension of the scholarship, and researchers themselves should be encouraged and incentivised only to be involved in activities that authentically reflect both the research and their research agenda. That principle guides our thinking. What flows from this is that we should aim to empower researchers by providing them with the mechanisms and encouragement to develop, and articulate their own narrative of engagement beyond academia.

---

<sup>45</sup> Trisha Greenhalgh et al, ‘Research impact: a narrative review’ (2016) 14(78) *BMC Medicine*, 14.

<sup>46</sup> Australian Research Council, *EI Submission Guidelines* (2018) 12.

Reframing the questions we ask ourselves can be powerful. So we propose that, rather than asking researchers to describe ‘what impact has your work had’, we move towards asking ‘*what have you done about your discovery?*’. While ‘all evaluation mechanisms distort the processes they purport to evaluate’,<sup>47</sup> we think the latter question would create fewer negative distortions than the former. The goal of the question is to move away from an externally-defined concept of ‘what impact is and how it is proved’ (an approach that encourages people to think in terms of standard metrics, like judicial citations or dollars saved or invested), and towards a reflexive approach where the researcher defines both what they were (or are) aiming to do for people, culture, society or the economy, *and how they have sought to achieve those goals.*

It refocuses the enquiry on translation activities that are within the researcher’s control, including but by no means limited to involving stakeholders in research from the start, giving public lectures or academic conference presentations, writing articles for blogs or news media or creating other accessible content (like podcasts), making data accessible to stakeholders, creating reports or pamphlets to disseminate key findings to practitioner or industry audiences, providing training, or arranging meetings and workshops with policymakers. Sometimes it will be possible to show that those activities have resulted in additional uptake or direct change, but the change of emphasis to activities *within the researcher’s control* and that are susceptible to be affected by existing privileges or advantages, and less reliant on networks of influence, working in well-populated fields and so on would promote healthier research cultures. We acknowledge that even this approach still risks privileging people who have the time and wherewithal to promote their own research. Giving presentations and creating reports is easier for those without caring responsibilities, and easier for researchers not overburdened with large teaching loads. We do however think that a reflexive approach, where the researcher defines what actions will best serve their goals, is at least *better*. In the end, no measure of achievement short of genuine engagement with performance relative to opportunity can address basic inequalities in time and space for research and research-related activity.

If impact - or as we would prefer, translation - questions were to be reframed in this way, researchers would be less disadvantaged for working in areas where achieving actual policy change is difficult, takes a long time, or where connections between research and impact elude documentation because they are too attenuated or non-linear. It would mean fewer incentives for researchers to overclaim their own individual responsibility for change, give less priority to the kind of doctrinal research that courts are most likely to cite, and make the substantial financial and career rewards that follow from appointments, promotions and grants less dependent on luck and privilege.

It may well be possible for law schools to do much to shift attitudes towards appointments, promotions and grants by making this adjustment internally, but some assessments - especially external assessments like the Engagement and Impact Assessment or equivalent

---

<sup>47</sup> Sir Gareth Roberts, *Review of research assessment* (2003) [‘Executive Summary’: 3] <[http://www.ra-review.ac.uk/reports/roberts/roberts\\_summary.pdf](http://www.ra-review.ac.uk/reports/roberts/roberts_summary.pdf)>, cited in Simon Smith, Vicky Ward, Allan House, ‘Impact’ in the proposals for the UK’s Research Excellence Framework: Shifting the boundaries of academic autonomy’ (2011) 40 *Research Policy* 1369, 1371.

government-directed assessment - will undoubtedly remain firmly fixed on demanding evidence of ‘actual’ impact. For these, we suggest that, as a discipline, we ought to become much more rigorous in demanding evidence that goes beyond proxies towards the attributes they seek to measure. For example, instead of simply reporting their number of citations, academics should be encouraged to explain their quality, such as the extent to which the citer has seriously grappled with the researcher’s evidence or reasoning. That is, we want to normalise researchers explaining *why* their citations are significant. Instead of simply reporting the prestige of the journals in which they publish, we should be asking researchers to provide more narrative explanation of why they published there.<sup>48</sup>

Similarly, insisting that the highest quality research be shoehorned into a small number of ‘prestige’ journals not only threatens the diversity of work produced, but it maintains those journals’ dominance, making it much harder, for example, to raise the quality and impact of emerging open access competitors (despite the academy’s urgent need to do so). Equally, if researchers are trying to reach specific audiences (for example industry, via practitioner-focused journals or grey literature) to maximise the reach and influence of their work, that should be rewarded too. We acknowledge that any new CALD ranking list will cut across our argument here. Nevertheless, we would argue that one method for pushing back against simplistic reliance on quantitative journal rankings is by encouraging academics to frame journal choices as part of a translation narrative that can sit alongside any story about the prestige of an academic’s publications.

## Mentoring and leadership

A second area of vital importance to the research enterprise, and which both universities and funders have shown increasing interest in assessing, is mentoring, leadership, collaboration and supervision. In our view, this collection of concepts represents fundamental values that foster research culture, and we agree that they should be promoted and rewarded. All too often the importance of academics’ contributions to the research environment is sidelined in individualistic processes like appointment, performance appraisal and promotion, or in impact assessments that put the emphasis on ‘rockstar’ academics. We should be encouraging researchers to see supporting the careers of others as a core part of the role, and trying to find ways to reward those who do it well. For this reason, and consistent with the reflexive approach we’re advocating, we think that it is important to be asking academics, as part of their overall research agenda narrative, to reflect on what they have done and will do to enable others’ research and research careers.

Here too, however, we face risks if we allow assessment of these contributions to be defined by easily measured proxies. The temptation is obvious: in grant applications, for example, especially at the most prestigious end (such as Laureate Fellowships), “look how many PhDs/ECRs I have pushed through the pipeline!” could easily become the simple way to

---

<sup>48</sup> If researchers were more thoughtful about journal choice, and required to give some kind of explanation, this could even have a salutary effect in encouraging more transparency from journals about their processes: to explain why publication in a particular journal *is* a marker of quality. It is, nevertheless, challenging for a researcher to access such information.

‘prove’ mentoring and leadership, and the convenient and ‘objective’ way for assessors and panels to compare candidates.

A numbers game in this area would be, we think, particularly pernicious because it risks providing incentives for behaviour that actively damages research environments and new careers. Mentoring has an evil twin: the kind of predatory academic ‘leadership’ which takes advantage of, or free-rides on, the work of (usually junior) students and colleagues. Examples include taking the benefit of being named as a co-author on a publication, or co-investigator on a grant without substantially contributing towards the work. It is perhaps obvious, but worth stating, that having more collaborators, or co-authors, or PhD students is not necessarily an indicator of the quality of those relationships. Indeed, one researcher having numerous research higher degree students might actively *detract* from the experience of those students given the supervisor’s finite time and attention is spread across a larger number of ‘beneficiaries’. This would be especially true in fields like law where it is relatively more rare for students to be working in laboratory-like settings where there are other sources of support such as other research students, postdocs and junior academics working on related projects. It is probably true that *some* legal researchers are moving towards a more lab-like model: all we are saying is that in the field of law, and probably many others, high numbers of RHD students or co-authors is not necessarily an indicator of a good collaborator and mentor, and cannot on its own distinguish between the good and evil twins. Information beyond mere numbers is required to make that judgment. In case our reader doubts that anyone would treat co-authorship as a proxy for ‘mentoring’, we note *Nature Communications* recently published an article which did precisely that.<sup>49</sup>

A second reason to doubt the use of numbers is based on the nature of law as a discipline. The last thing we want to do is encourage or allow the uncritical use of proxies that could disadvantage our discipline when (simplistically) compared to others. Law has a strong tradition of the solo researcher and sole authorship, even in cases where researchers are collaborating via other means (such as reading and commenting on others’ work). Even among top researchers and in the most well-regarded journals in Australia and globally, a minority of research papers have more than one author.<sup>50</sup> Similarly, reliance on numbers of PhD students would favour a subset of specialisations within legal research. Although we are not aware of any attempt to measure this empirically, our anecdotal impression is that demand from students to undertake PhDs varies significantly and systematically across the discipline of law, likely because holding a PhD is not necessary for aspiring commercial

---

<sup>49</sup> Bedoor AlShebli, Kinga Makovi, and Talal Rahwan, ‘The Association between Early Career Informal Mentorship in Academic Collaborations and Junior Author Performance’ (2020) 11(1) *Nature Communications* 5855. The authors of this questionable study even went a step further to assert that they had evidence that female researchers mentored by other women were disadvantaged compared to those mentored by men, claiming that: ‘having more female mentors is associated with a decrease in the mentorship outcome, and this decrease can reach as high as 35%, depending on the number of mentors and the proportion of female mentors’. Use of the proxy in the paper attracted instant negative comments from reviewers and readers.

<sup>50</sup> Ian Murray and Natalie Skead, “‘Who Publishes Where?’: Who publishes in Australia’s top Law journals and which Australians publish in top Global Law journals?” (2020) 47(2) *University of Western Australia Law Review* 220, 242 [confirm proportion]. Interestingly, Murray and Skead note (at 243) that 60% of their top 50 authors have one or more co-authored articles in leading law journals; but (perhaps unlike US) the rise in co-authorship does not appear to be driven by empirical or interdisciplinary work, perhaps because that work is still less welcome at generalist law journals than it should be.

lawyers and barristers (and because the opportunity costs of that further study are so high). This is not just a risk, but a reality: we are aware of one Australian law faculty which set research supervision targets that are simply unachievable for academics working in most legal disciplines (a ‘minimum’ of 6 students for Level Es, regardless of subdiscipline!). This makes it impossible for academics to hit their baseline targets, and adds yet more pressure to perform elsewhere. The push for higher supervisions also creates perverse incentives for academics to compete for the best students, regardless of whether they are in fact the best supervisor for the project - or the reallocation of PhD students across faculty in a way that disadvantages students by placing them with inappropriate or inexperienced supervisors.

In short, the usual numerical proxies that are used to indicate collaboration or mentorship in some other academic disciplines (however badly) are seriously problematic for law. More generally, numbers alone give us little information about whether a researcher is a *good* mentor, leader or collaborator. Ensuring that rewards are reserved for *good, balanced* collaboration and not damaging forms depends critically, we think, on more qualitative information about what the researcher is bringing to those collaborations. That information could come from the researcher themselves articulating their role, and from peer and expert assessment of claims.

### Asking the right question: what have you done to enable others?

Just as we believe that research translation is critical, so too we think that creating a collaborative, enabling research culture is a key part of the role of a research academic that we need to foster and reward. Direct research collaboration in the form of joint projects or publications won't be appropriate for every field or every question, but still, if we are going to tackle the really big, really difficult questions, and do so without unnecessary waste and duplication, we need researchers to be prepared to work together and communicate better - because no single discipline or subdiscipline will have all the answers, and because we need to work faster to solve our most pressing problems, from pandemics, to social inequality, to environmental collapse.

We think that one way to start a conversation about leadership, collaboration and mentoring is to encourage researchers to reflect on, and articulate, '*what have you done to enable the research and careers of others?*'. Asking researchers, and assessors, to think about how their research has enabled others focuses attention on some of the obvious markers of leadership frequently recognised by various university awards for leadership and mentoring, such as whether researchers are involved in capacity-building, HDR supervision, and mentoring of junior researchers and peers; creating opportunities for others such as by leading major grant projects, and/or including scholarships and junior appointments in research projects. But our choice of the word 'enabling' (rather than 'leadership') is very deliberate. Traditional academic leadership isn't for everyone, and isn't the only valuable way to enable others.

As with our discussion of translation above, we would be aiming to encourage researchers to develop their own narrative, and to see enabling others as part of the research enterprise and agenda. Thinking this way opens up a range of possibilities beyond direct collaboration. Sharing resources (such as datasets) is one, as is publishing research in open access

formats.<sup>51</sup> Actively contributing as part of a research team on a project developed by others is too, because it's vital to getting the work done. Constructive peer review and assessment is another way that researchers can enable others' careers, and can potentially be proved through recognition in authors' acknowledgments. So is using opportunities to bring good research by less well-known researchers to broader attention. Part of the lasting influence of Professor Jill McKeough has been via her role in the establishment of the Australasian IP Academics' conference and her ongoing promotion of its extraordinarily welcoming culture that supports junior scholars in the field. The inclusion of this question encourages academics to consider what they can do to disrupt the university as a 'privilege machine'<sup>52</sup> - indeed, which explicitly makes it part of their job to do so.

Another way that researchers could show how they have enabled others is by making special efforts to include more diverse voices in scholarship. This is important, not only to disrupt existing privilege and assist disadvantaged groups overcome hurdles to their success, but because the inclusion of diverse voices enriches institutions and scholarship, and brings to light perspectives and experiences that ought to inform how we develop the law and regulatory institutions. This is perhaps especially important for law schools, which train significant proportions of society's future leaders. Academics could, for example, evidence efforts of this kind by providing analyses of whom they cite: pointing out, for example, their efforts to cite women, scholars of colour, and scholars from less privileged institutions and countries, or include them in workshops and edited collections. Even being aware of the risk of gender bias in citation practices, and documenting efforts to counter that bias, could benefit a researcher's own scholarship (by requiring them to widen their own searches), benefit underrepresented groups (by introducing their work to broader audiences) and the field of scholarship (by widening the perspectives reflected in the work). More broadly, academics could explain where they have gone beyond ordinary expectations to help recruit/support junior researchers and create pathways to permanent positions for people who are, for example, people of colour, have a first language other than English, have a disability, or are first in family to go to university. We hasten to add that we are not advocating for cringeworthy saviour-complex behaviour, or charity, and we recognise the risk that even including this as a suggestion risks precisely that. Solutions to avoiding this aren't easy, but we think that the reflexive approach assists here too: perhaps researchers could be encouraged to evidence how they enable others while (1) respecting and promoting their autonomy (for example, not constantly suggesting that indigenous scholars work on indigenous legal issues or sit on indigenous committees); (2) recognising their achievements, potential and value; (3) never exploiting any power imbalances in the relationship; (4) actively seeking to level the playing field.

Asking this question of how the researcher has enabled the research and careers of others also interacts in interesting and worthwhile ways with our earlier discussion of impact and what a researcher has done about their discoveries. Discussing together how one's work *enables others* as well as how it has had some effect in the world could promote a thoughtful

---

<sup>51</sup> Professor Katy Barnett generously shared her painstakingly-constructed and coded dataset of High Court citations with us for the purpose of preparing this paper - just one example of an enabling research behaviour.

<sup>52</sup> Raewyn Connell, *The Good University: What Universities Actually Do and Why It's Time for Radical Change* (Bloomsbury Publishing, 2019), chapter 5.

engagement with impact that acknowledges the reality that it is “often difficult to attribute the opportunity to make an impact to any given piece of research”.<sup>53</sup> Rather than demanding researchers claim the whole of some change in the world to themselves, they could frame change as a collaborative effort and highlight the ways their work has enabled progress in the field together with others.

## What’s needed to implement such a framework in practice?

The questions we pose above invite us to focus our attention more squarely on what we’re doing and why it’s important, both in articulating our own research agenda and plans, and assessing the research and broader performance of others. But if as a discipline we agree that we want people to ask these questions, and create their own research narratives that articulate how the answers to these questions fit into their research agendas and projects, then together we will need to provide researchers with the space and the reasons to do so - *without* creating yet another administrative burden or check box exercise. In this chapter, we are inviting reflection as to how, as a community, we can empower ourselves to better promote what we think is valuable about the broader enterprise of university research in law. We want to build a ‘healthy microbiome’ which supports the assessment of research and researchers on impact, translation, mentoring and leadership and which resists colonisation by weird metrics and inappropriate proxies from outside. But how?

### 1. Awareness-raising and collective action

Cliche’d as it might seem, raising understanding across the discipline of the problems we have highlighted around research assessment is, we think, the first step. The first time that many academics (including us) engage seriously with how their research will be assessed, or how they can tell a story that clearly articulates what research and translation they are doing, and how it all fits together, is when they apply for promotion or a grant. Not only is this often too late - retro-fitted narratives are never as convincing - but it also makes it more likely that the only kind of evidence people will reach for is whatever numbers they’ve been told about (and seen cited in institutional precedents). We freely admit that, even though we are both experienced researchers and research assessors, and one of us has even been an Associate Dean Research, we didn’t appreciate the full extent of the use of metrics, or the empirical work demonstrating the many problems with proxies, until we started systematically investigating them for this chapter. We strongly suspect we are not alone. And since the meta-research literature is large, there are doubtless other considerations and problems with assessment proxies we haven’t mentioned.

So: we need to talk. This means not just private conversations between Associates Dean Research, or the occasional senior scholarly voice in an interview or promotion panel, and not just at the country’s most prestigious and privileged law schools. *All* legal researchers need to be aware of the pitfalls of celebrating High Court (or even judicial) citations as the pinnacle of legal research ‘impact’, or believing that ‘research translation’ means ‘law reform’. We need to be better empowering *all* legal researchers to articulate their own research, translation and collaboration as a coherent, research-driven story.

---

<sup>53</sup> Simon Smith, Vicky Ward, Allan House, ‘Impact’ in the proposals for the UK’s Research Excellence Framework: Shifting the boundaries of academic autonomy’ (2011) 40 *Research Policy* 1369, 1376.

There are obvious ways to further these conversations. We can run (plenary) sessions in specialist legal conferences dedicated to professional development: sessions of this kind have long been part of the IP Academics' conferences which Jill McKeough began. We could also create tools for researchers: for example combining the various insights about data and proxies in this chapter (and beyond) to create useful, credible, informative documents that could be used by both applicants, and assessors, in writing or assessing applications for grants or promotions etc, to avoid expectations that are unspoken or unwritten. Research performance criteria, and ARC grants commonly ask applicants to describe the impact of their research. As a discipline - for example via CALD, or within individual/groups of law schools - we could provide law researchers and their assessors (within and outside law) with explanations of what impact or translation means for us; the problem with demanding linear stories of how impact happens, and give them a range of ideas for how they can plan, describe, and evidence 'what they have done about their discovery'. We could also explain the weaknesses of commonly used proxies (like citations) with reference to the research. Similarly on mentoring and leadership, we could provide the text for applicants (or assessors on multi-disciplinary assessment committees) to explain the nature of collaboration in legal research, challenge common metrics (like numbers of RHD students or co-authorship), and again give applicants some ideas about how to think about how their work enables others.

An important role of these kinds of tools, text or guides is that, when backed up with startling data like that of Barnett on the gendered nature of High Court citations, they make it so much harder for committees of academics - whether for hiring, promotion, or grant allocation - to fall back on the data. We know that, both as applicants *and* as reviewers, we can be short of time and risk-averse, which can lead to 'conservative and narrow' conceptions of both impact<sup>54</sup> and collaboration. And we know too that committees feel a lot more secure comparing numbers than forming judgements about, and comparing, the importance or meaningfulness of candidates' work.<sup>55</sup> Part of what we need to do, then, is to make this less comfortable: i.e. to make it difficult or impossible for assessors to rely on inappropriate metrics, by confronting them with both evidence of their problems, and alternatives.

Guides of this kind would be stronger if they reflected a discipline-level consensus, developed cross-institutionally or via CALD. The tendency of tools like this would hopefully be to strengthen, and spread across the discipline, what we already see in practice. The legal academy has already pushed back by assessing impact broadly when it comes to ERA impact assessments, where broad interpretations were adopted and many case studies were highly ranked, and in assessing ARC grant applications, where the 2010 journal rankings list was pretty unceremoniously dumped, and h-indexes are disdained. As we noted at the outset, academics largely audit themselves, so they should use the opportunity to articulate their own research values and help the discipline as a whole to do so.

---

<sup>54</sup> Even those research assessment frameworks that appear neutral on their face, can end up being interpreted narrowly: Simon Smith, Vicky Ward, Allan House, 'Impact' in the proposals for the UK's Research Excellence Framework: Shifting the boundaries of academic autonomy' (2011) 40 *Research Policy* 1369, 1371; Derrick, *The Evaluators' Eye: Impact Assessment and Academic Peer Review* (Palgrave Macmillan, 2018).

<sup>55</sup> Mark W J Ferguson, 'Treat metrics only as surrogates', (27 October 2016) 538 *Nature* 455.

## 2. Social accountability

Changing a culture is not straightforward, so it is worth consulting the research we have on analogous cultural change. One area to which we could draw an analogy is diversity in institutions such as workplaces. The empirical studies in this area show there's lots of things that *don't* work to promote diversity - people forget their diversity training within days, and 'a number of studies suggest that it can activate bias or spark a backlash'.<sup>56</sup> Hiring tests tend to be used and interpreted selectively, and can end up actually reducing the representation of women and people of colour in senior ranks. In other words, merely *telling* people the cultural change required isn't going to achieve it. Simply telling people about the importance of collaboration and mentoring is unlikely to change much either.

Something that *has* been shown to work, however, is *social accountability*. This tactic 'plays on our need to look good in the eyes of those around us.' Experiments have shown that, when people discover their decisions may be reviewed by people whose opinions they care about, it helps them make better judgments (whether it's about the grade their students should receive, or raises for workers) and bias all but disappears.<sup>57</sup> This leads us to wonder whether there may be a role for assessment of assessors. For example, what if ARC reviewers were themselves given a letter grade of A-E based on the appropriateness of their reviews? The Colleges are no doubt informally making such a judgment in deciding how much weight to give them - feeding that judgment back may be a useful reminder to assessors that their assessments are being evaluated, creating a similar leveling effect. Some journals already do this: the *Indian Law Review*, for example, gives editors the ability to assign scores to reviewers based on the helpfulness of their review, although at present that is not fed back directly to reviewers. But what if they were? Could this simultaneously provide incentives for better reviews - *and* give reviewers the evidence they need to show how they have enabled the research and careers of others?

## 3. Systems adjustment and alignment

Collective action and awareness, however, are unlikely to be sufficient on their own to change research culture or empower legal researchers to think differently about what they are doing about their discovery, or how they are enabling others. It would be particularly problematic to try to 'raise awareness' by creating yet another (inconsistent, additional) reporting system. The last thing we want to do is generate even *more* audit culture that diverts even more precious research time. Building, writing, presenting, evaluating, prioritising, and selecting CVs is a prolific and time-consuming industry for grant applicants, faculty candidates and assessment committees alike.<sup>58</sup> 'Compliance-creep' - the time required to comply with auditing - leads to continually less time for the work we're trying to measure actually getting done. A particular challenge in what we have been arguing for is that we are advocating for a nuanced and holistic narrative of a researcher's agenda. Narratives of this kind create work: both for researchers engaged in writing them up and

---

<sup>56</sup> Frank Dobbin and Alexandra Kalev, 'Why Diversity Programs Fail', *Harvard Business Review* (July - August 2016) <<https://hbr.org/2016/07/why-diversity-programs-fail>>

<sup>57</sup> Frank Dobbin and Alexandra Kalev, 'Why Diversity Programs Fail', *Harvard Business Review* (July - August 2016) <<https://hbr.org/2016/07/why-diversity-programs-fail>>.

<sup>58</sup> David Moher et al, 'Assessing scientists for hiring, promotion, and tenure' (2018) 16(3) *PLoS Biology* e2004089.

evidencing any claims, and for assessors. We have already noted that doing this properly requires space: not overly-constraining word limits that force oversimplification and the removal of nuance. Unless researchers are given good reasons or incentives to engage in this process, this isn't going to happen in any meaningful way. As Connell has noted, 'evading intrusive and time-wasting management requirements has become a necessary skill for university workers.'<sup>59</sup>

As we noted at the outset, researchers are already asked to articulate their research agenda, the quality of their research, its impact, and their contributions to leadership and mentoring in a range of identifiable circumstances: the key ones being when they (1) apply for a job, (2) report annual performance, (3) apply for promotion and (4) apply for (internal or external) research funding, and in some cases, (5) articulate a case study for the purposes of an external impact assessment. These requirements - although ultimately all concerned with the same basic goal of assessing research quality - are often framed differently, with different criteria. For example, Monash University has very different criteria for promotion than for appointment at the same level - *even though it's the exact same job*. ARC grant and ERA criteria are different again from university appointment, performance, or promotion criteria. Annual performance systems ask for different information from that required for promotion applications. And so on.

Now imagine if these systems asked the *same* questions? Then researchers could use the same material, or hand copies of material to professional staff tasked with writing up university-level reports such as the ERA. We could 'write it once, and properly' rather than constantly writing up new narratives to meet slightly different questions. We hasten to add that this does *not* mean we are advocating for a single system into which academics would be required to input their articulation of their research agenda as we have broadly defined it, that would then automatically populate other digital systems. It will frequently be the case that researchers will wish to highlight or emphasise particular aspects of their research career in showing why they are the appropriate person to be involved in, for example, a grant application. No researcher's story will be so one-dimensional that it can be described, once and for all. What we are advocating however is that the systems and the various reporting obligations should be broadly consistent, and ask the same kinds of questions, so that researchers can at least re-use and tweak material rather than rewrite from scratch on a regular basis, and so it's clearer to emerging researchers what's expected of them from the get-go.

## Conclusions

For the most part, in law at least, it is the most senior academics holding permanent positions who are asked to do most of the work in assessing research impact, as well as appointments, promotions, performance appraisals and grant applications. That puts the most privileged of us in positions to shape these research assessments in ways that promote the values and behaviours we wish to see rewarded. That invites the question - what are the research goals and values that really matter? And how do we help create the

---

<sup>59</sup> Raewyn Connell, *The Good University: What Universities Actually Do and Why It's Time for Radical Change* (Bloomsbury Publishing, 2019) 133, footnote omitted.

incentives, and the space, for researchers to do great research responsibly and well? In this chapter we have suggested some ways to reframe concepts that we are *already* starting to be assessed on (impact, mentorship etc), to make sure that we ask the right questions for our discipline. Dispiriting as the current moment in higher education is, and inexorable as the tide of auditing may seem, research impact assessment done *right* can be empowering to the individual academic framing their own agenda in these terms, their institution and the discipline - and encouraging researchers to think about what they are doing, and why. Similarly, encouraging researchers to more actively enable the research and careers of others can help us not only best address way-of-life-threatening challenges to life, climate and governance, but foster a healthy, supportive environment in which to do so.