

Pontika, N., Klebel, T., Correia, A., Metzler, H., Knoth, P., and Ross-Hellauer, T. (2022) Indicators of research quality, quantity, openness and responsibility in institutional review, promotion and tenure policies across seven countries . Quantitative Science Studies. Advance Publication. https://doi.org/10.1162/qss_a_00224

Indicators of research quality, quantity, openness and responsibility in institutional review, promotion and tenure policies across seven countries

AUTHORS

Nancy Pontika, Open University, UK. nancy.pontika@open.ac.uk, <https://orcid.org/0000-0002-2091-0402>

Thomas Klebel, Know-Center GmbH, Austria. tklebel@know-center.at, <https://orcid.org/0000-0002-7331-4751>

Antonia Correia, University of Minho, Portugal. antonia.correia@usdb.uminho.pt, <https://orcid.org/0000-0002-6610-8853>

Hannah Metzler, Medical University of Vienna, Complexity Science Hub Vienna and Institute for Globally Distributed Open Research and Education, Austria. metzler@cs.h.ac.at, <https://orcid.org/0000-0001-9254-3675>

Petr Knoth, Open University, UK. petr.knoth@open.ac.uk, <https://orcid.org/0000-0003-1161-7359>

Tony Ross-Hellauer, Know-Center GmbH and TU Graz, Austria. tross@know-center.at, <https://orcid.org/0000-0003-4470-7027>¹

ABSTRACT

The need to reform research assessment processes related to career advancement at research institutions has become increasingly recognised in recent years, especially to better foster open and responsible research practices. Current assessment criteria are believed to focus too heavily on inappropriate criteria related to productivity and quantity as opposed to quality, collaborative open research practices, and the socio-economic impact of research. Evidence of the extent of these issues is urgently needed to inform actions for reform, however. We analyse current practices as revealed by documentation on institutional review, promotion and tenure processes in seven countries (Austria, Brazil, Germany, India, Portugal, United Kingdom and United States of America). Through systematic coding and analysis of 143 RPT

¹ Corresponding author, Open and Reproducible Research Group, Graz University of Technology and Know-Center, Inffeldgasse 13, 8010 Graz, Austria, <https://orrg.eu/>

policy documents from 107 institutions for the prevalence of 17 criteria (including those related to qualitative or quantitative assessment of research, service to the institution or profession, and open and responsible research practices), we compare assessment practices across a range of international institutions to significantly broaden this evidence-base. Although prevalence of indicators varies considerably between countries, overall we find that currently open and responsible research practices are minimally rewarded and problematic practices of quantification continue to dominate.

KEYWORDS

Open science, Responsible Research and Innovation, rewards and recognition, research assessment, indicators.

1 INTRODUCTION

The need to reform research assessment processes related to career advancement at research institutions has become increasingly recognised in recent years, especially to better foster open and responsible research practices². In particular, it is claimed that current practices focus too much on quantitative measures over qualitative measures (Malsch 2015; Colavizza et al. 2020), with misuse of quantitative research metrics, including the Journal Impact Factor, among the most pressing issues for equitable research assessment generally, and to foster open and responsible research in particular. Therefore, recent years have seen a focus on attempts to understand how principles and practices of openness and responsibility are currently valued in the reward and incentive structures of research performing organisations, especially by direct examination of organisations’ review, promotion and tenure³ (henceforth RPT) policies. Such studies have heretofore focused on specific contexts, however. While work led by Erin McKiernan and Juan Pablo Alperin examined policies in place across a range of types of institutions in the US and Canada (Alperin et al. 2019; Alperin et al. 2020; McKiernan et al.

² We use this broad term in this paper to denote the confluence of two sometimes overlapping trends in research reform: (1) Open Science, “the movement to make scientific research, data and dissemination accessible to all levels of an inquiring society” including Open Access to publications, data-sharing and increased collaboration in research (Pontika et al. 2015). Responsible Research and Innovation is an umbrella concept, dominant in Europe, which denotes research which is made responsive to society through practices like public engagement, open access, gender equality, science education, ethics and governance (Owen, Macnaghten, and Stilgoe 2012; von Schomberg 2019).

³ The systems of promotion and tenure differ across these countries. While tenure and procedures to award it are central in the USA, German-speaking countries still largely rely on a “chair-system”, with a stark division between full professors and non-professorial staff (Brechelmacher et al. 2015). Here, we analyse all assessment processes for progression in academic careers, including all variants of tenure.

2016; Niles et al. 2020), Rice et al. (2020) studied criteria used across a range of countries, but only within biomedical sciences faculties. Hence, further work is needed to describe types of criteria in place across a range of institutional types internationally.

This paper aims to fill this gap. Our primary research question can be formulated: “What quantitative and qualitative criteria for review, promotion and tenure are in use across research institutions in a purposive sample of seven countries internationally?” Sub-questions include: “How prevalent are criteria related to open and responsible research in these contexts?”, “How prevalent are potentially problematic practices (e.g., use of publication quantity or journal impact factors)?”, and “What trends can be observed across this sample?”.

In order to answer these questions, we investigate the prevalence of qualitative and quantitative indicators in RPT policies across seven countries: Austria, Brazil, Germany, India, Portugal, United Kingdom and United States of America. This involved manually collecting 143 RPT policy documents from 107 institutions. These documents were then systematically coded for the inclusion of language related to 17⁴ elements (including those related to qualitative or quantitative assessment of research, service to the institution or profession, and open and responsible research practices) using a predefined data-charting form. Directly comparing the

⁴ We initially coded the additional indicator “general impact”, but discarded it for the analysis since it largely reflected an aggregate measure of other indicators on impact beyond academia (public, industry, policy-making).

indicators and criteria in place at such a range of international institutions hence aims to broaden the evidence-base of the range of practices currently in place.⁵

2 BACKGROUND

2.1 Research assessment and researcher motivations

Institutional policies regarding RPT typically focus on three broad areas: research, teaching, and service (both to the profession and the institution). The relative importance of each varies across institutions and has also changed over time (Gardner and Veliz 2014; Youn and Price 2009). In the European context, a recent survey of researchers investigated indicators widely used at EU institutions for review, promotion and tenure. The most common factors used in research assessment were (according to survey respondents): number of publications (68%), patents and securing funds (35%), teaching activities (34%), collaboration with other researchers (32%), collaboration with industry (26%), participation in scientific conferences (31%), supervision of young researchers (25%), awards (23%) and contribution to institutional visibility (17%) (European Commission. Directorate General for Research and Innovation. 2017).

When it comes to the assessment of research contributions, reflecting the common idiom “publish or perish”, publication in peer-reviewed venues remains central. Primary publication types vary across disciplines. While journal articles dominate in Science, Technology,

⁵ Results from this study were previously made available via the ON-MERRIT project report “D6.1 Investigating Institutional Structures of Reward & Recognition in Open Science & RRI” (Pontika et al. 2021). This paper presents enhanced analysis based on a slightly modified dataset (corrected to eliminate minor inconsistencies in data-charting, as explained in footnote 8, methods section). In addition, the data underlying this study is also incorporated into the data paper “How do career promotion policies affect research publications and open access? [version 1; peer review: awaiting peer review]” (Pontika et al. 2022a).

Engineering and Mathematics (STEM) subjects, monographs or edited collections have greater importance in the Humanities and Social Sciences (Adler, Ewing, and Taylor 2009; Alperin et al. 2020). Within Computer Science, meanwhile, publication in conference proceedings is the most important factor (McGill and Settle 2011). However, irrespective of which type of publication is favoured, institutions tend to position *productivity* (often quantified via metrics) as a defining feature in RPT policies (Gardner and Veliz 2014). The ways in which this emphasis on productivity and quantification influences academics' focus and shapes behaviours, often in detrimental ways, is worth expanding upon in order to understand how current trends in RPT policies may be limiting uptake of open and responsible research.

Institutional committees tasked with determining whether research contributions are sufficient for promotion, review or tenure, face something of a dilemma. While the quantity of publications is comparatively easy to assess, measuring their quality is a more difficult challenge. Ideally committees would be able to read each of the contributions themselves to make their own first-hand judgements on the matter. However, the mass of material created, as well as increased research specialisation drastically reducing the number of experts that possess the required expertise for such quality judgements, mean that usually proxy indicators for quality are sought. Here, two factors are particularly popular: publication venue and citation counts.

In perceptions of the prestige of academic journals, the Journal Impact Factor has assumed a particularly pernicious role. Created by Eugene Garfield of the Institute for Scientific Information, the Journal Impact Factor calculates an average of citations per article within the

last two years to provide a metric of the relative use of academic literature at the journal level. Originally created to assist library decisions regarding journal subscriptions, the Journal Impact Factor soon came to be used as a proxy for relative journal importance by research assessors and researchers themselves (Adler, Ewing, and Taylor 2009; Walker et al. 2010). Various criticisms have been levelled at the Journal Impact Factor, most prominently that relatively few outlier publications with many citations skew distributions such that most publications in that journal fall far below the mean. Additional criticisms include that differences in citation practices between (and even within) fields make the Journal Impact Factor a poor tool for comparison, that it is susceptible to gaming by questionable editorial practices, and suffers a lack of transparency and reproducibility (Fleck 2013). Nonetheless, its use as a proxy for research quality in research assessment became commonplace (Gardner and Veliz 2014; McKiernan et al. 2019). McKiernan et al. (2019) studied RPT documents and found that 40% of North American research-intensive institutions mentioned the Journal Impact Factor or closely-related terms. Accordingly, researchers commonly list a journal's Impact Factor as a key factor they take into account when deciding where to publish (Niles et al. 2020).

Citation counts at the article level are also often used as a proxy for research quality within RPT processes (Adler, Ewing, and Taylor 2009; Brown 2014). Indeed, Alperin et al. (2019) found that such indicators were mentioned by the vast majority of institutions. However, citations have been widely criticised for being too narrow a measure of research quality (Wilsdon et al. 2015; Hicks et al. 2015; Curry 2018). The application of particularistic standards is especially perilous for early-career researchers who have yet to build their profile. By using citation metrics to

evaluate research contributions, initial positive feedback leads to the self-reinforcement loop known as the Matthew effect (Wang 2014). Moreover, indicators such as the h-index are highly reactive (Fleck 2013) and therefore risk reifying monopolisation of resources (prestige, recognition, money) in the hands of a select elite. The h-index was designed as a measurement tool to showcase the consistency of the cited researchers, but creates a disadvantage for early career researchers and neglects the diversity of citation rates across scientific disciplines and subdisciplines (Costas and Bordons 2007).

2.2 Research assessment and open and reproducible research

Multiple initiatives in the last decade have sought to raise the alarm on the overuse of quantitative indicators and highlight the need to consider a broader range of practices (beyond publications). For instance, the ‘San Francisco Declaration on Research Assessment’ (DORA), specifically criticised use of the Journal Impact Factor in research assessment.⁶ The ten principles of the ‘Leiden Manifesto for Research Metrics’ (Hicks et al. 2015) sought to reorient the use of metrics by critiquing their “misplaced concreteness and false precision”, arguing that quantitative should be used as a support for “qualitative, expert assessment”, with strict commitments to transparency.

Such critiques of over-quantification have developed alongside movements to foster open and reproducible research. These two trends meet where advocates of Open Science or Responsible Research & Innovation (RRI) identify concern amongst researchers that uptake of open and

⁶ San Francisco Declaration of Research Assessment: <https://sfdora.org/>

responsible research practices will negatively impact their career progress (Adler, Ewing, and Taylor 2009; Migheli and Ramello 2014; Rodriguez 2014; Peekhaus and Proferes 2015; Wilsdon et al. 2015).

As a result, recent research has investigated if and how criteria relating to open and responsible research practices are rewarded in RPT policies. In particular, the influential “Promotion, Review and Tenure” project headed by Erin McKiernan and Juan Pablo Alperin has examined these issues in depth by studying a corpus of RPT policy documents from 129 universities in the USA and Canada. This project found that aspects related to open and reproducible research were rare or undervalued. Alperin et al. (2019) found, for example, that only 6% of RPT policies mentioned “Open Access”, often in a negative way. Public engagement, although mentioned in a large number of policies, was nonetheless undervalued by associating it with service, rather than research work. Meanwhile, 40% of policies from research-intensive institutions mentioned the Journal Impact Factor in some way, with the overwhelming majority of those (87%) supporting its use in at least one RPT document and none heavily criticising it (McKiernan et al. 2019).

A similar study by Rice et al. (2020) studied the presence of “traditional” (e.g., publication quantity) and “non-traditional” (e.g., data-sharing) criteria used for promotion and tenure in biomedical sciences faculties. In that context, the authors found that mention of practices associated with open research were very rare (data-sharing in just 1%, with Open Access publishing, registering research, and adherence to reporting guidelines mentioned in none). Most prevalent were traditional criteria including “peer reviewed publications” (95%), grant funding

(67%), “national or international reputation” (48%), “authorship order” (37%), journal impact factor (28%) and citations (26%).

While general trends, including prevalence of (sometimes problematic) quantitative measures and lack of recognition for open and responsible research practices, can be observed across these two groups of work, nonetheless there are important nuances we should take into account. In the biomedical context, Rice et al. (2020) saw “notable differences” in the availability of guideline documents across continents and “subtle differences in the use of specific criteria” across countries. In the US/Canada context, meanwhile, differences across types of institutions were observed - with, for instance, “research-intensive” institutions being more likely to encourage use of the Journal Impact Factor (McKiernan et al. 2019). These differences, across institutional types and national boundaries, require further investigation.

This current study complements and extends this work. Such work is crucially important especially since reform of rewards and recognition processes is now a policy priority, especially in Europe. Vanguard institutions such as Utrecht University in the Netherlands are already implementing such reforms (Woolston 2021). The Paris Call on Research Assessment, announced at the Paris Open Science European Conference (organised by the French Presidency of the Council of the European Union) in February 2022, calls for evaluating the “*full range of research outputs in all their diversity and evaluating them on their intrinsic merits and impact*” (“Paris Call on Research Assessment” 2022). The Paris Call also sought the formation of a “coalition of the willing” to build consensus and momentum across institutions. The European

Commission is currently building such a coalition (Research and Innovation 2022). This paper further contributes to the evidence-base to inform such reform.

3 METHODS

We assembled and qualitatively analysed review, promotion and tenure policy documents (RPT) from academic institutions in seven countries (Austria, Brazil, Germany, India, Portugal, UK, USA).

3.1 Sampling

In selecting countries, we used purposive sampling (candidate countries whose primary language was covered by the research team - i.e., English, German or Portuguese). While automated translations can go a long way in basic understanding, the task at hand required knowledge of the policy landscape of the studied countries, as well as the ability to precise reading of source materials. We first identified four target countries, based on our European focus and our team's familiarity with the language (English, German, Portuguese) and policy landscapes of specific countries. In addition, the USA was included as a representative of a leading research country and to allow comparisons with previous research (Alperin et al. 2019). Further, we included India and Brazil as examples of large "low and middle income countries" based on gross national income per capita as published by the World Bank.⁷ They play a growing role in research, and broaden our scope to include Asia and South America. We acknowledge that our sample of countries cannot be considered random or representative of the situation globally. However,

⁷ World Bank <https://www.worldbank.org/en/home>

given the current lack of knowledge of RPT criteria in place across national contexts, we nonetheless believe that our sample adds richly to current knowledge.

To include representative numbers of institutions of perceived high- and low-prestige, we used the Times Higher Education World University Rankings (WUR) 2020 to select institutions. Institutions from each selected country were sorted based on their relative WUR performances in the categories “Research” and “Citations”. We then divided each category into three equally sized subcategories, defining them as “High-”, “Medium-” and “Low-” performing institutions. Next, we calculated the median of each subcategory and selected the institutions that were closest to the median as representatives of this category. We included both the “Research” and “Citations” fields from the WUR since both of these are research-related indicators. Duplicate entries of the same institution appearing in both categories were replaced by the next available institution in the “Citations” category. Our sampling procedure resulted in a sample of 107 institutions across 7 countries (Table 1).

Table 1. Number of institutions per country - the total number of institutions sampled relative to the number of institutions listed per country in WUR

Country	# of institutions in WUR (2020)	# of institutions in sample	% of sampled institutions	# of policies analysed
Austria	11	6	54.5%	13

Brazil	46	12	26.1%	15
Germany	48	12	25.0%	13
India	56	12	21.4%	12
Portugal	13	6	46.2%	7
UK	100	24	24.0%	47
USA	172	35	20.3%	36
TOTAL	446	107	24.0%	143

While selection of institutions based on university rankings is an often-used strategy (e.g., Rice et al. 2020 use the Leiden Ranking in a similar approach), it is not without flaws. First, university rankings have been criticised for their reliance on biased and unreliable reputational survey data (Waltman et al. 2012) and issues of gaming and selective reporting (Gadd 2021). Second, rankings such as the THE WUR only include the most prominent institutions, and leave out many institutions based on partly arbitrary criteria (e.g., how many yearly publications they need to be included). The reported groups of “high”, “medium” and “low”-ranked universities are only relative to the set of institutions included in the ranking, and not academia as a whole. We see our use of the WUR as a pragmatic approach to reproducibly sampling universities. This does not negate their deficiencies for guiding prospective students to choose institutions or informing policy decisions.

3.2 Data Collection

Policy documents were collected using a shared search protocol. First, we used Google to search for the institution name along with various constellations of keywords. Table 2 shows the set of keywords identified and used for the policies identification in the three languages: English, German and Portuguese.

Table 2. Search key terms in English, German and Portuguese which were used to retrieve related RPT policies

English	German	Portuguese
Policy	Satzung, Richtlinie, Verfahren	Política, Regulamento
Review	Qualifikationsprüfung, Review, Beurteilung, Leistungsevaluation, Regelung, Richtlinie, Strategie	revisão
Academic, Researcher, Professor	wissenschaftliche Mitarbeiter, (Junior-) Professor	académico, universitário, investigador, professor
Promotion	Beförderung, Promotion, Berufung	promoção

Institutions often have both institution-wide and departmental-specific RPT policies. Due to difficulties in finding specific departmental policies in the UK and USA, we collected only institution-level policies. To ensure a consistent set of policies, we defined the following exclusion rules:

1. We did not collect advertisements for job descriptions even though these could include some insightful requirements applicable to the RPT policies.

2. We included RPT policies only and not other policies such as Ethics, Diversity, OA, where similar concepts could appear.

Policies could apply to any post-PhD researcher career stage. The collected policies evaluated various research-related positions. For example, in the UK, some institutions have separate policies for 1) associate professors, 2) full professors, and 3) readers. In the USA, there are separate policies for 1) tenured and 2) non-tenured staff. In Austria, there are policies for 1) habilitation (qualification for teaching, needed for promotion to professor) and 2) qualification agreements for tenure track (associate professors), while no promotion to full professors exists. In India, we could often not find specific policies, but rather the evaluation forms that researchers use to apply for promotion. In these cases, we therefore analyse the evaluation forms instead.

Some institutions have separate policies for all researcher categories, (i.e., separate policies for lecturers, assistant professors, associate professors, professors, and so on), while others have a uniform policy covering all positions. Hence, the number of institutions is smaller than the total number of policies collected (Table 1). Where more than one policy was identified for an institution, we assessed the indicators separately for each policy and counted an indicator as “fulfilled” when it appeared in at least one policy.

We were sometimes unable to obtain policy documents for target institutions. Specifically, where access was restricted to members of the institution only, data collectors emailed the institution’s

human resources department to ask for a copy of the policy. If no response was received within ten days, data collectors recorded this information and sampled the next institution from the list for that country and strata until a sufficient number of policies was obtained. Table 3 below shows the institutions that did not have a public policy per country.

Table 3. Total number of institutions we checked with no policy

Country	# of institutions with no publicly available policy	total # of institutions checked
Austria	4	10
Brazil	8	20
Germany	5	17
India ⁸	27	39
Portugal	0	6
UK	32	56
USA	6	41
TOTAL	82	189

⁸ For the Indian case, most institutions did not have specific policies. To retain India in the sample, we therefore coded 5 evaluation forms, 2 policy documents and 5 documents that included both a policy and an evaluation form.

An initial round of document collection occurred during the period November 2019 to March 2020. The sample was then further extended between March and April 2021. As some institutions had several distinct policies relating to different career-stages, 143 total RPT policy documents were collected for analysis from the 107 institutions.

3.3 Data charting

Data was extracted from the policy documents using a standardised data-charting form. The form was devised in multiple rounds of iteration. Key indicators for inclusion were identified from various sources including the MoRRI indicators (MoRRI, 2018), a group of studies performed in the North American context by Alperin et al. (2019), as well as from the surveyed literature. We collected and examined 17 different indicators (Table 4), including "traditional" assessment indicators relating to quantification and quality of publications, and a set of "alternative" indicators relating to open and responsible research, and related issues such as gender equality and Citizen Science. In addition, information was gathered on the date policies came into effect, the academic positions (e.g., tenure track, professor, lecturer, senior lecturer) or types of processes (e.g., promotion, review, tenure) they governed.

Table 4. Overview of data-charting form main elements ("Are the following mentioned as being taken into account in the promotion/evaluation procedures as stated in the policy?")

Indicator	Short code	Definition / coder instructions
Citations	Citations	Mentions citations (h-index, number of citations etc) (Y/N)
Citizen Science	CitizenScience (SL3)	Mentions citizen science activities (research activities involving

		citizens incl. data collection, agenda-setting, etc.) (Y/N)
Data	Data (OA2)	Mentions data publication and/or sharing (Y/N)
Engagement with industry	Industry	Mentions impact type 3) interactions with industry (Y/N)
Engagement with policy makers	PolicyMakers (PE2)	Mentions impact type 2) interactions with policy makers (Y/N)
Engagement with the public	PublicEngagement (PE1)	Mentions impact type 1) public engagement (policy specifically mentions public engagement, e.g. science communication, public lectures, social media, radio, newspapers) (Y/N)
Gender equality	GenderEquality (GE1)	Mentions gender Overview or equivalent process (gender of the candidate is considered) (Y/N)
Journal metrics	JournalMetrics	Mentions journal based metrics (e.g. impact factor, journal ranking quartiles) (Y/N)
Number of publications	NumPublications	Mentions number of publications as an indicator of productivity in and of itself Y/N) ⁹
Open Access	OpenAccess (OA1)	Mentions Open Access publishing (Y/N)
Patents	Patents	Mentions patenting discoveries/inventions (Y/N)
Publication quality	PublicationQuality	Mentions an attempt to review quality instead of quantity of publications, i.e. a request for key publications, writing samples,

⁹ The indicator was initially broader in scope, also covering cases mentioning a full list of publications as a mandatory part of the documents to be submitted. As we elaborate in the discussion, these requests might imply that the number of publications is being considered, but this is not evident from the policy itself. We therefore revised the indicator, referring strictly to instances where the number of publications was taken as an indicator of productivity in and of itself (e.g., “numbers help”, University of Nevada, Las Vegas, USA_31) and recoded documents accordingly.

		etc. (Y/N)
Review & editorial activities	PeerReview	Mentions peer review and/or editorial activities
Reviewer gender	GenderReviewers	Gender of reviewers considered (e.g. reviewers must include women but requirement of gender parity not mentioned) (Y/N)
Reviewer gender balance	GenderBalanceReviewers (GE9)	Mentions gender balance for reviewers/committees as a requirement/goal (Y/N)
Service to profession	PastoralWork	Mentions conference organisation, participation in PhD committees or other community services (Y/N)
Software	Software	Mentions contributions to scientific software (Y/N)

Five coders were involved, all with competences in English, three in German and one in Portuguese. Policies were assigned based on language competences. They coded the presence (1) or absence (0) of each indicator in each policy, and copied the sentence mentioning the indicator, and the ones before and after. Each document was coded by one individual. To assess intercoder reliability, an independent coder (TRH) performed a reviewer audit of a random sample of ten percent of the total number of institutions. Comparing this second round of review to the first responses revealed a high intercoder-reliability of 96.78%.

Before carrying out the analysis, several steps were taken to ensure data integrity and consistency. Data was originally collected via spreadsheets, and subsequently collated using R to avoid copy-paste errors. We checked (a) that every indicator was present for each policy, (b) that in cases where an indicator had been found (coded as 1) a text excerpt was present, and that (c)

in cases where no indicator had been found (coded as 0) also no text excerpt was present. The inconsistencies found were checked and resolved by TK.

To facilitate the review of our results, and the re-use of the data, we translated all non-English excerpts to English in a two-step procedure. First, we used DeepL¹⁰ to obtain an initial translation of the excerpt. A native speaker then checked the translation, revising to ensure that meaning, context and use of special terms mirrored the original. The validated translation was then recorded alongside the original text for subsequent analysis.

3.4 Data Analysis

All data analysis was conducted using R (R Core Team 2021), with the aid of many packages from the *tidyverse* (Wickham et al. 2019), including *ggplot2* for visualisations (Wickham 2016). Computational reproducibility of the analysis is ensured through the use of the *drake* package (Landau 2018). The analyses presented in this paper are all exploratory and have not been pre-registered. To enable the comparison of the indicators' presence against each other, we rely on Multiple Correspondence Analysis (M. Greenacre and Nenadic 2018; Nenadic and Greenacre 2007). Correspondence analysis and its extension MCA are similar to principal component analysis (PCA) in mapping the relationships between variables to a high-dimensional Euclidean space. The goal of the method is then “to redefine the dimensions of the space so that the principal dimensions capture the most variance possible, allowing for lower-dimensional descriptions of the data” (Blasius and Greenacre 2006, 5). The obtained dimensions can therefore

¹⁰ <https://www.deepl.com/translator>

be inspected for their alignment to specific variables, enabling conclusions about the main trends found in the data. MCA thus offers a visual representation of contingency tables, and is well suited for the categorical data collected in this study. Furthermore, MCA allows us to investigate the relationship between indicators and countries jointly. Several considerations apply when analysing data via MCA.

First, we apply MCA in a strictly exploratory fashion. Inspecting its visual output facilitates interpretation of the relationship between indicators, and how they relate to countries, but we do not conduct any testing of hypotheses. Second, the graphical solution offered by MCA maximises deviations from the average, allowing for statements of the prevalence of indicators relative to one another. For statements about absolute frequencies of indicators across countries, we rely on MCA's numerical output (see supplementary files), as well as cell frequencies found in the corresponding contingency tables. All supporting data and required code are available via Zenodo (Pontika et al 2022b).

4 RESULTS

Assessing the prevalence of traditional and alternative (especially open/responsible research-related) criteria across policies of 107 institutions, we find substantial differences in their prevalence (Figure 1). While 72% of institutions mention “service to the profession”, no institution mentions data sharing or Open Access publishing. Overall, traditional indicators,

related to the profession or to scientific publications, are much more common than indicators related to open and responsible research.

In terms of more traditional indicators, by far the most common indicator mentioned in the policies was “service to the profession”, which includes activities such as organising conferences or mentoring PhDs (72%). Extending the concept of professional service, almost half of the policies also mention peer review & editorial activities (47%). A second important aspect among the sampled policies is that of scientific publications, with frequent mentions of the number of publications, or publication quality. While a call to rate quality over quantity is not uncommon in the policies, problematic practices were still worryingly prevalent. For example, journal metrics such as the Journal Impact Factor were mentioned in at least a quarter of the policies, while sheer productivity, as measured by quantity of publications, was present in around a fifth of cases.

Indicators relating to open and responsible research were very rare. We discovered no mentions of data-sharing or Open Access publishing. Creation of software was quite well represented (13% of cases), due to its prevalence in policies in Brazil where it is mentioned at 75% of institutions (Figure 3). Mentions of RRI elements were more encouraging, as the RRI-related aspects of interactions with industry (37%), engagement with the public (35%), and engagement with policy-makers (22%) were relatively well-represented. However, issues relating to gender were mentioned only in between 6 to 9% of cases.

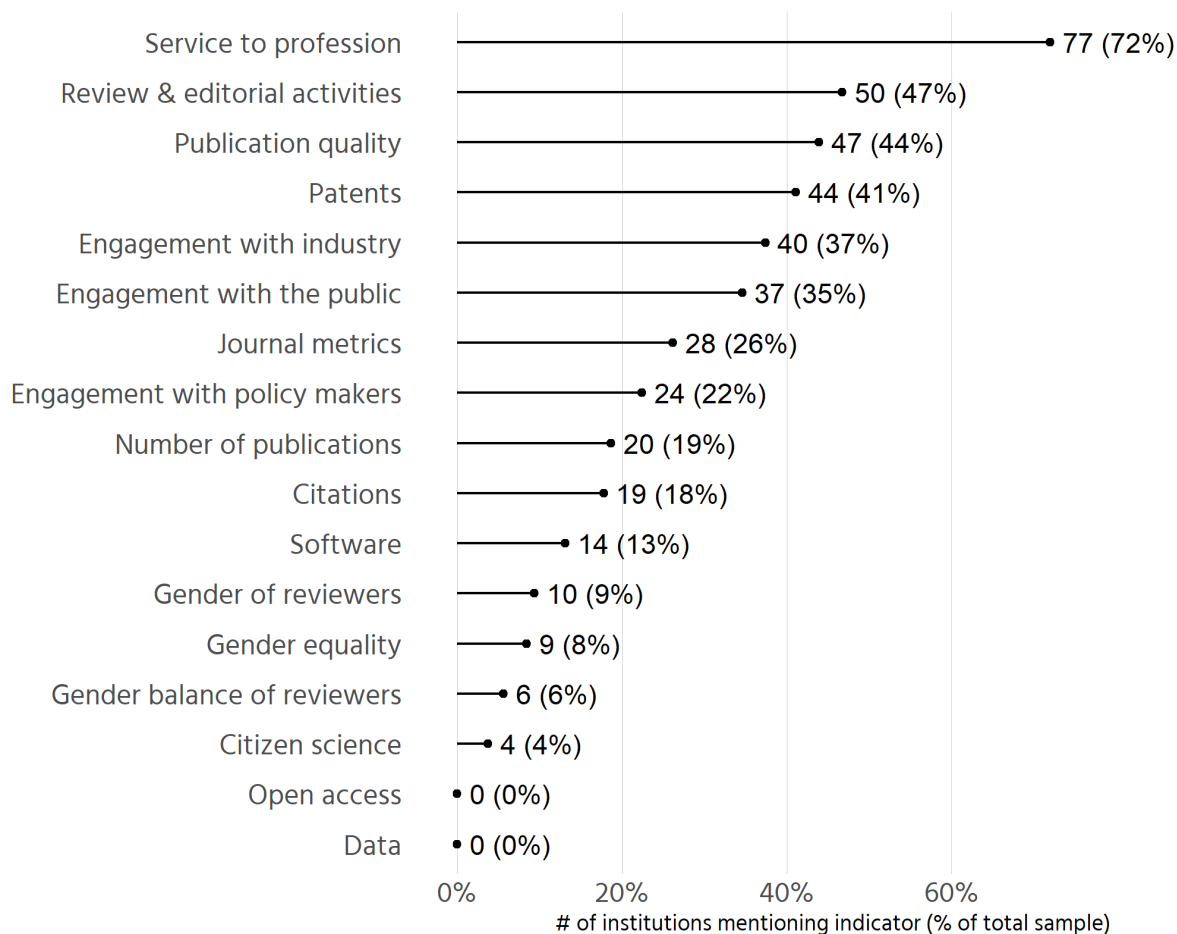


Figure 1. Overall prevalence of indicators across all institutions/countries. In cases where an institution had more than one policy, we aggregated the policies. An institution was counted as mentioning a given indicator, if at least one of the policies mentioned it.

4.1 Relationship between indicators

Institutions rely on a distinct combination of indicators and criteria to assess researchers. To investigate how these indicators are related (i.e., which aspects are commonly mentioned in tandem), we rely on Multiple Correspondence Analysis. This method relates criteria against each

other and allows us to investigate deviations from the average, as well as which indicators commonly appear together. To further substantiate these findings, we provide bivariate correlations between all indicators in the supplement (Figure S2). Note that these analyses are exploratory, and based on a dataset of moderate size.

The first apparent aspect from analysing the variables jointly is that the studied indicators tend to be cumulative (Figure 2). In broad terms, there is a basic divide between institutions that mention many criteria, and those that mention few to none (see also Table 5 on how this relates to countries). Investigating relationships further, the first dimension (horizontal axis) draws heavily from engagement beyond academia: with industry, the public, and policy makers. Institutions commonly mention them together, with bivariate correlations of about .5 between the three indicators (see Figure S2). The same institutions also mention contributions to review & editorial activities, service to the profession and publication quality more often than the average institution. On the other end of the spectrum (right side) are institutions that mention engagement beyond academia and service to the profession less frequently than the average.

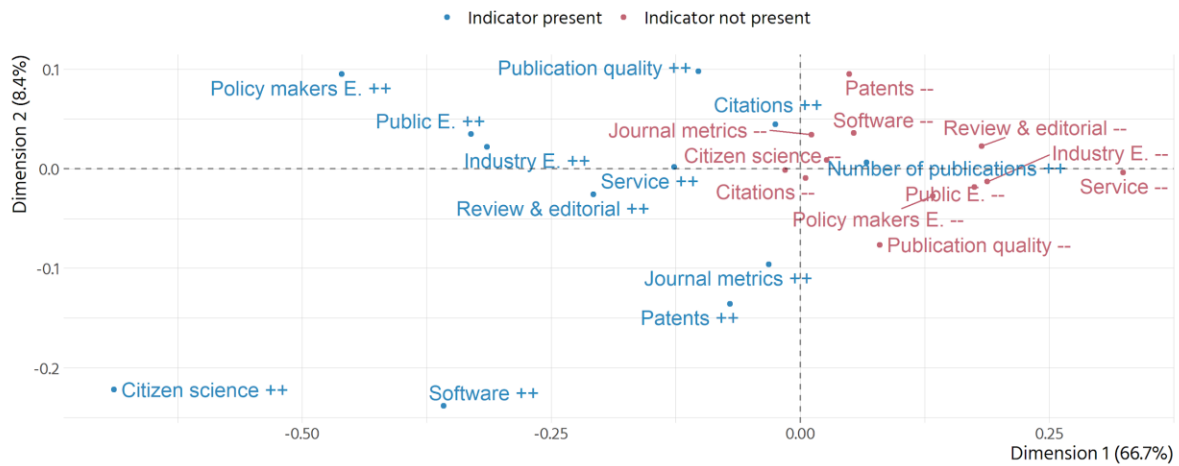


Figure 2: Relationship between indicators for review, promotion and tenure. The figure is a graphical representation of the relationships between indicators when considering their multivariate relationships. The figure's origin (0, 0) represents the sample average. “++” means that an indicator is present, “--” that it is not present. Engagement is abbreviated with “E.” The *horizontal axis (Dimension 1)* accounts for 66.7% of variation in the data. This dimension mainly contrasts institutions which mention instances of engagement (with the public, industry, or policy makers), citizen science, as well as service to the profession, with institutions that do neither. The *vertical axis (Dimension 2)* accounts for 8.4% of variation in the data. This dimension mainly contrasts institutions that value publication quality and rely on citations, with institutions that value patents, journal metrics, as well as software on the other end of the spectrum (bottom). Citizen science is found near the bottom of the axis, but does not contribute strongly to this dimension. The indicators “Data sharing” and “Open Access publishing” are not included in the model, since both were not found in any of the policies. Furthermore, variables relating to gender were not included, since they relate to the composition of review panels rather than research assessment criteria per se.

While the first distinction (between institutions mentioning many indicators and engagement beyond academia in particular) is strongest, the second dimension (vertical) provides additional insight on the interrelatedness of the indicators. The divergence along this dimension revolves around institutions which mention publication quality and reliance on citations on one side, with institutions mentioning software, patents, journal metrics, as well as citizen science, on the other side. It is noteworthy that publication quality and citations (often seen as a proxy indicator for publication quality), are mentioned jointly at an above average rate. Mentions of publication quality, on the other hand, are unrelated to mentions of journal metrics (such as the Journal Impact Factor, $r = -0.01$, 95% basic bootstrap CI $[-0.21, 0.17]$), which are considered a much more problematic indicator of research quality. Citations and journal metrics are represented on opposite sides of the vertical spectrum, despite the criteria being weakly correlated ($r = .22$, $[0.02, 0.44]$). This is driven by the fact that journal metrics are moderately related to patents ($r = .32$, $[0.14, 0.52]$), but unrelated to publication quality. It should be noted that the concepts of “journal metrics” and “publication quality” might overlap in how they are applied in practice. While we coded text phrases such as “High quality scholarly outputs with significant authorship contributions.” as pertaining to publication quality, this might in practice be assessed via journal metrics (e.g., Journal Impact Factor or ranking quartiles).

4.2 Country comparison

When comparing countries, we find differences in terms of the overall prevalence of indicators, but also their relative importance. The absolute number of indicators per country varies since we sampled more institutions for larger countries than for smaller ones. Importantly, however, the relative number of indicators also varies considerably (Table 5). While just under a third of the

analysed indicators were identified in policies in Austria, Brazil, Germany, Portugal and the UK, 18% in the USA, and 16% in India. The lower numbers in the USA and India may reflect the nature of the documents examined in those cases. We only examined institution-wide policies and in the USA it may be the case that detailed criteria are more often contained at departmental- or faculty-level policies; in India (as stated) assessment forms were also analysed as few institutions had official policy documents (see Table 3).

Table 5. Number and percentage of indicators discovered per country

Country	# of universities	# of indicators	# of indicators found	% of all possible indicators
Austria	6	17	29	28%
Brazil	12	17	61	30%
Germany	12	17	64	31%
UK	24	17	109	27%
India	12	17	33	16%
Portugal	6	17	27	26%
USA	35	17	106	18%

Regarding the relative prevalence of the various indicators, we find considerable differences between countries. Figures 3 & 4 rely on the same data, and enable a thorough examination of the differences present in our data across countries. In the following we summarise results for each country in detail.

- Austria: A very high share of sampled institutions mention the number of publications (67%), and half also mention journal metrics. Service to the profession, while the most common concept across countries, is mentioned in only 50% of institutions in Austria. A major distinction between institutions from German speaking countries (i.e., Austria and Germany) and all other countries, is that the former frequently mention concepts of gender, with four out of six Austrian universities mentioning gender equality, while this is not found in any other country.
- Brazil: All Brazilian institutions mention service to the profession, and three out of four mention patents, review & editorial activities, and software, while mentions of software are uncommon in other countries. Similar to India and Austria, journal metrics are mentioned quite frequently (42%). Sampled policies from Brazil are similar to policies from the UK in frequently mentioning service and engagement beyond academic, but diametrically opposed in also frequently mentioning patents and software, both of which are very rare in the UK. Finally, both country profiles are relatively far from the sample average, indicating configurations which are less common among other countries.
- Germany: Policies from German universities are very similar to their Austrian counterparts, which suggests similarities based on shared cultural and academic traditions and influences. For example, policies from both Austria and Germany commonly mention gender equity. However, the concepts of patenting and review & editorial activities appear considerably more frequently in German policies than in Austria, with the least mentions of service to the profession across the sample also found in Germany.

- India: Contrary to all other countries, we find no evidence of policies referring to review & editorial activities as a criterion for promotion, and very few cases that refer to the number of publications a given researcher has produced. On the other hand, mentions of journal metrics were very common in the policies sampled from Indian institutions (67%, n=8), while less common among the other countries.
- Portugal: All sampled universities mention engagement with the public, which is a strong exception in the sample. Furthermore, many institutional policies mention service to the profession, engagement with industry, patents, as well as the number of publications. Indicators which are less common across the sample, such as citizen science, software, citations, as well as engagement with policy makers are not found at all in Portugal.
- United Kingdom: All sampled universities mention service to the profession, and four fifths mention publication quality. Equally, institutions from the UK mention all three dimensions of engagement beyond academia (industry, public, policy makers) considerably more frequently than the average of the sample. Finally, policies mention patents and the number of publications considerably less frequently than institutions from other countries.
- United States: In line with the overall finding of a low propensity of indicators across universities from the US (Table 5), all indicators were found at a slightly lower rate than in other countries. Given that the sample deliberately included more institutions from the US, institutional policies from the US are quite close to the average across the whole sample (Figure 4). The biggest deviations from the sample average are found with

engagement with industry, which is mentioned least frequently in the US, compared to all other countries.

Service to profession	50%	100%	33%	58%	83%	100%	63%
Patents	33%	75%	67%	67%	67%	4%	34%
Review & editorial activities	17%	75%	75%	0%	50%	58%	40%
Engagement with industry	33%	33%	33%	25%	83%	62%	20%
Engagement with the public	17%	42%	25%	8%	100%	62%	17%
Publication quality	33%	0%	58%	33%	17%	79%	40%
Journal metrics	50%	42%	25%	67%	17%	12%	14%
Number of publications	67%	25%	25%	8%	33%	4%	17%
Engagement with policy makers	17%	33%	8%	0%	0%	54%	14%
Gender of reviewers	50%	0%	58%	0%	0%	0%	0%
Gender equality	67%	0%	42%	0%	0%	0%	0%
Citations	17%	0%	33%	8%	0%	17%	26%
Software	0%	75%	8%	0%	0%	0%	11%
Gender balance of reviewers	33%	0%	33%	0%	0%	0%	0%
Citizen science	0%	8%	8%	0%	0%	0%	6%
Open access	0%	0%	0%	0%	0%	0%	0%
Data	0%	0%	0%	0%	0%	0%	0%
	Austria	Brazil	Germany	India	Portugal	United Kingdom	United States

Figure 3. Prevalence of indicators per country. Percentages are rounded to full integers. The number of cases (universities) per country is presented in Table 5. Low cell frequencies and empty cells prohibit the use of common Chi-square metrics for contingency tables.

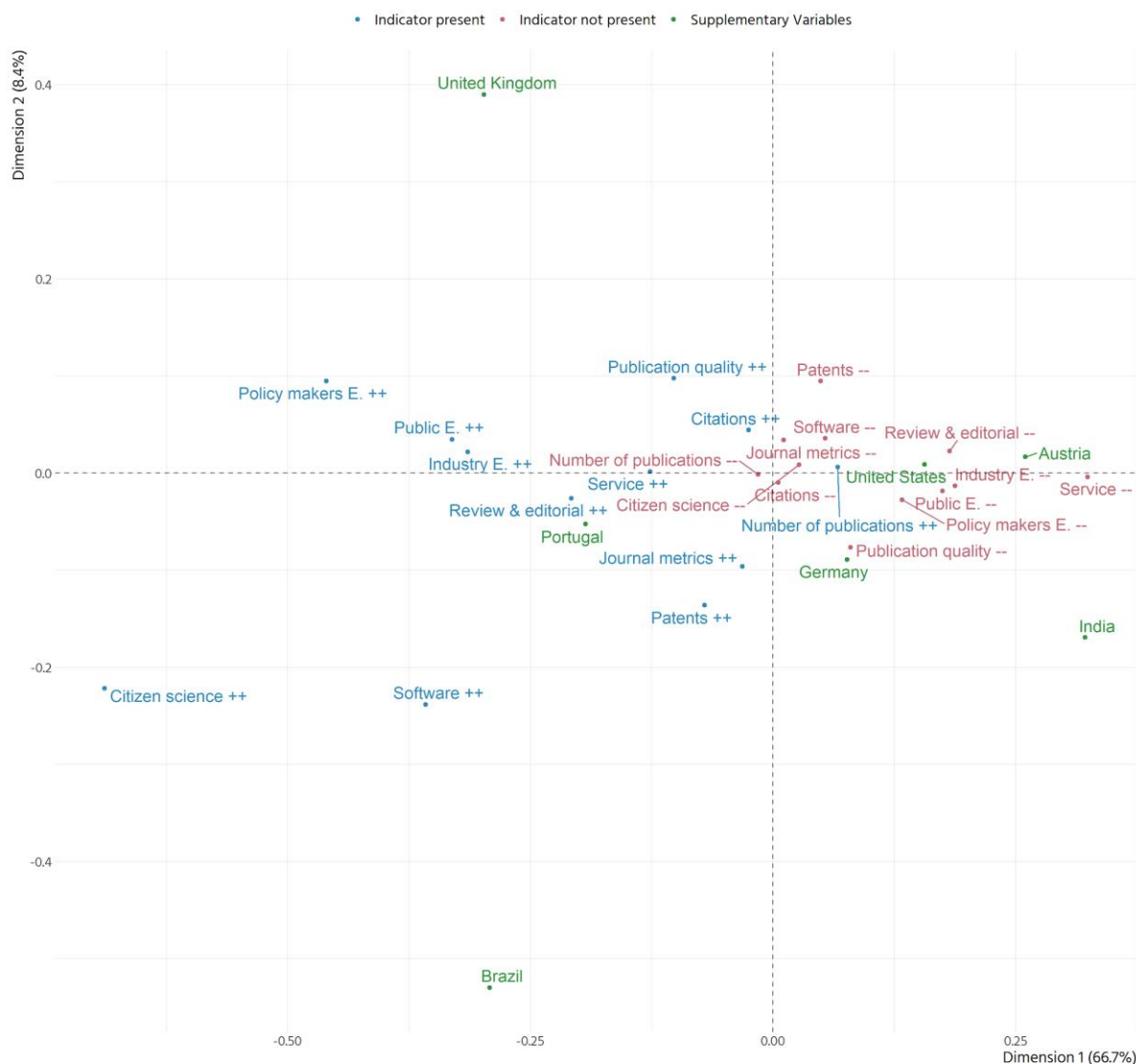


Figure 4: Relationship between indicators with superimposed countries. The relationships between criteria displayed in this figure are the same as above (Figure 3). To allow for an investigation of which criteria are more common in a given country than in the rest of the sample, we project country profiles into this space. These

“supplementary variables” do not have an influence on the layout of the indicators. The countries’ positions are to be interpreted as projections onto the respective axes, by examining their distance to indicators which are central to the respective dimension (see section 3.1 and Figure 3 on the interpretation of the axes).

Overall, we find a low uptake of alternative evaluation criteria covering open and responsible research. However, there is substantial variation between countries (Figure 5). Summarising eight core criteria (“Citizen science”, “Data”, “Engagement with industry”, “Engagement with policy makers”, “Engagement with the public”, “Gender equality”, “Open Access”, “Software”), we find the highest uptake of alternative criteria in Brazil, Portugal and the UK, with 1.9, 1.8 and 1.8 alternative criteria per university on average, followed by the UK (1.8). Uptake is lower in Austria and Germany, and particularly low in the USA (0.7 criteria on average) and India (.3).

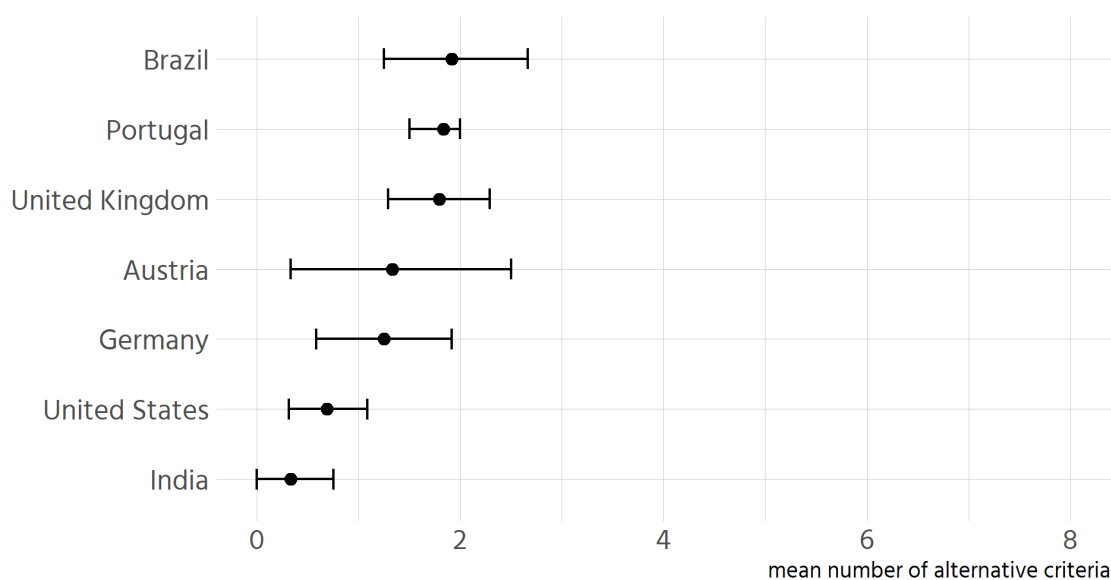


Figure 5: Uptake of alternative indicators. Here we display how frequently alternative indicators are found in the policies. We consider the following eight indicators: “Citizen science”, “Data”, “Engagement with industry”,

“Engagement with policy makers”, “Engagement with the public”, “Gender equality”, “Open Access”, “Software”. Dots represent the mean across all universities of a given country, with bootstrapped confidence intervals (95%).

4.3 Comparison with citation ranking

Previous research has found no evidence of an association between universities’ ranking positions and the prevalence of traditional or alternative criteria when controlling for geographic region (Rice et al. 2020). Here we conduct a similar analysis, examining differences in the citation ranking and its relationship to the set of criteria, while controlling for country. Removing the influence of countries is meaningful in this context, because an institution's location and its citation ranking are clearly linked (Figure S4).

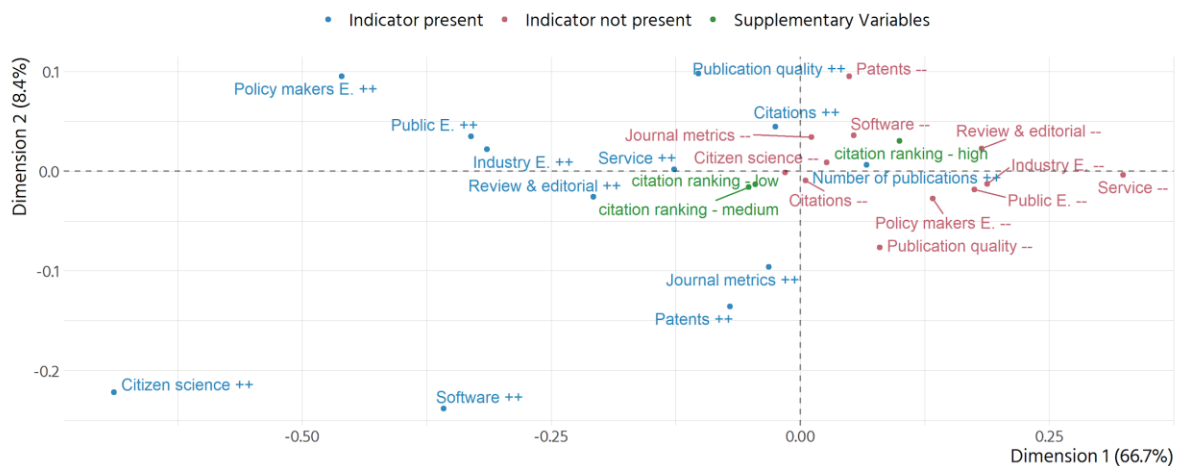


Figure 6: Relationship between indicators with superimposed citation ranking groups. The relationships between criteria displayed in this figure are the same as above (Figure 3). To allow for an investigation of which criteria are more common in a given ranking group than in the rest of the sample, we project the respective

profiles into this space. These “supplementary variables” do not have an influence on the layout of the indicators. Ranking categories are calculated within-country, to control for the influence of country on an institution's citation ranking. The ranking positions are to be interpreted as projections onto the respective axes, by examining their distance to indicators which are central to the respective dimension (see section 3.1 and Figure 3 on the interpretation of the axes).

After controlling for country, we find only small differences in the prevalence of indicators with respect to an institution's ranking (Figure 6). Institutions with a low as well as with a medium citation ranking are very close to the sample average on both dimensions. Both are characterised by slightly above average mentions of the dimension of engagement beyond academia, as well as the dimension of service. Highly ranked institutions are characterised by slightly lower than average mentions of service, review & editorial activities, and engagement, but slightly higher mentions of publication quality, citations, and journal metrics (see also Figure S5).

4.4 “Numbers help”? Journal metrics, publication quantities and publication quality

We next look further into the ways in which two problematic practices (use of journal-level metrics and numbers of publications as indicators of quality and productivity respectively) are expressed in policies, as well as how policies discuss publication quality.

More than a quarter of the policies we examined mention the Journal Impact Factor or some other measure of journal/venue prestige as an assumed proxy for the quality of research published there. This was highest in India (67%) and Austria (50%). In the latter, unambiguous use of the Journal Impact Factor was found: "The evaluation is based on the journal rankings according to the impact factors from the unchanged ranking lists of the Institute of Scientific

Information (ISI)" (Medical University of Vienna, AT_4a). Brazil (42%) also relied heavily on journal-level metrics, specifically the "QUALIS-CAPES classification" (Pinto, Matias, and González 2016). Use of such metrics was least visible in the UK, where the 14% of policies that mention them also tend to be more circumspect in their language, e.g., "Excellence might be evidenced [...] (in part) by proxies such as journal impact factors" (Teesside University, GB_6).

Numbers of publications as a criterion are present in around one in five policies, invoked in various ways. This criterion is especially common in Austria (67%), (e.g., "The list of publications of a habilitation candidate must include at least 16 scientific publications in international relevant journals with peer review procedures, which have been published in the last 12 years", Medical University of Innsbruck, AT_3). Such quantification is sometimes used in the contexts of strict formulas which also used journal-level metrics (as at the aforementioned Medical University of Vienna (AT_4a): "The basic requirement for a habilitation is 14 points, with 1 point for a standard paper and 2 points for a top paper"). In the USA, quantity of publications is mentioned in 17% of cases, but usually emphasised as just one factor amongst others, e.g., "Quantity can be a consideration but quality must be the primary one" (University of Missouri-St Louis, USA_25). In the striking words of one USA institution, however, "*numbers help*" [emphasis ours] when reporting "the total number of peer-review articles or other creative and research outputs" (University of Nevada, Las Vegas, USA_31). In Germany, only 25% of institutions focused on publication numbers, but often emphasising "not to set a fixed minimum number, but rather an approximate guideline" (TU Dortmund, DE_12). However, in Austria and Germany we also found that as a matter of course many institutions ask for full publication lists

as part of their criteria. We did not code these as explicitly supporting publication quantity as an indicator for assessment. However, in practice we might assume that the length of publication lists may be used as an unofficial factor in decisions.

As with journal metrics, UK policies only very rarely mention publication numbers as a factor (just 4%). This is in stark contrast to the number of UK policies mentioning publication quality as an important criterion (79%). Here, the influence of initiatives like DORA and the UK's Forum for Responsible Metrics, and the way these have translated into the UK national assessment exercise, the Research Excellence Framework (REF), is clearly visible. We found, for example, exhortations to produce "high quality" work "that is judged through peer review as being internationally excellent or better in terms of originality, significance and rigour" (University of Sheffield, GB_9a). As we discuss below, this language is highly similar to that of the REF itself, suggesting that institutions have adapted their assessment policies to REF criteria.

5 DISCUSSION

The need to reform reward and recognition structures for researchers to mitigate effects of over-quantification and incentivise uptake of open and responsible research practices is well understood. Our results show just how far there is to go.

We found that policies for assessing researchers members for review, promotion and tenure among an international sample of 107 institutions in seven countries largely relied on traditional criteria (service to professions, review & editorial activities, publication quality and patents).

Alternative criteria related to open and responsible practices were much less prevalent. Here, considerations related to Responsible Research and Innovation such as engagement with industry, the public and policy-makers fared better, present in between 22% and 37% of policies. Gender elements (including commitments to gender equity and gender balance of reviewers) were only present in between 6% and 9% of cases, and only found in Austria and Germany. Criteria related to Open Science were very rare - sharing of data and Open Access publishing were not found in any policy. These general findings across countries hence largely confirm previous findings which have focused on USA/Canadian institutions (Alperin et al. 2019; McKiernan et al. 2019; Alperin et al. 2020; Niles et al. 2020) or a particular discipline (Rice et al. 2020).

Regarding differences between the countries studied, we found substantial differences with some common patterns. Overall, we found very few criteria in the policies from India and the USA. India likely constitutes a special case, with institutions often seeming to lack written policies beyond those implied by the required criteria in application forms. For the case of the USA, the distinction between general RPT policies and those from specific departments and schools is crucial. The analysed policies represented general policies, which in many cases laid out general principles, but did not include more specific criteria, which were to be defined by each school or department. This is in contrast to policies from Austria or Germany, where the policies equally applied university-wide; however these policies were very specific and we did not find evidence of further policies at faculty- or department-level.

Criteria relating to engagement beyond academia (industry, policy-makers, public) appeared together very often (correlations around $r = 0.5$), and most commonly in Brazil, Portugal and the UK. The high share of UK institutions mentioning this type of outreach can be related to the influence of the Research Excellence Framework (REF) as an organising principle. 25% of the profile for an institutional score in the REF is attributed to “Impact”, defined as “an effect, change or benefit beyond academia” (Sutton 2020), and the prevalence of broader impact criteria in current institutional policies seems to reflect this importance. Similarly, the effect of the REF on definitions of quality of outputs and the diminished use of journal metrics as a proxy for quality can clearly be seen in the UK, where as journal metrics were barely mentioned and many policies focused on the quality of publications themselves was foregrounded, sometimes literally using the REF definition of “Quality of research in terms of originality, significance and rigour”, as in the case of (University of Sheffield, GB_9a) quoted above.

There are high similarities between the UK, Portugal and Brazil, with a strong emphasis on service, review & editorial activities, and the dimension of engagement beyond academia. However, the UK is very distinct from Portugal and Brazil in terms of patents and publication quality. While patents are very frequently mentioned in Brazil and Portugal, but almost not at all in the UK, publication quality was found very often in the UK but not at all in Brazil, and only in one out of six universities in Portugal.

Relationships between the presence or absence of specific indicators and an institution's relative level of “prestige” (imperfectly captured here via their citation ranking) are weak. Our findings

align with results by Rice et al. (2020), who reported statistically insignificant coefficients for universities' ranking positions on the uptake of traditional and alternative indicators, after controlling for country. Institutional policies explicitly rewarding high levels of citations or publications in journals with a high Journal Impact Factor do not seem to translate easily to an institution's increased success in this regard. Hence, not only are such policies problematic in incentivising gaming of metrics and potentially fostering bad practices (Ioannidis 2005; Higginson and Munafò 2016), but they do not even necessarily work as desired to raise an institution's position in such rankings.

In addition, that the prestige of institutions is less of a factor than we may have expected, spotlights the extent to which local or regional norms of research assessment must be further studied. We suggest that a main takeaway from our findings is that although the overall trends (predominance of traditional and quantitative indicators and general lack of newer metrics of open and reproducible practices) are visible across the countries, substantial differences in the emphasis on specific sets of criteria exist. This has implications for reform of reward and recognition.

Current RPT policies result from a complex network of factors, including diverging evaluation cultures, differing levels of institutional autonomy, and institutional preferences (Seipel 2003; Adler, Ewing, and Taylor 2009; Coonin and Younce 2009; King, Acord, and Earl-Novell, Sarah 2010; Walker et al. 2010; McGill and Settle 2011; Gardner and Veliz 2014; Brown 2014). The best route forward on reforming these local assessment cultures is thus to be decided in light of

historical and contextual considerations. Strict, one-size-fits-all reforms would be insufficient in bringing about desired outcomes across the complex web of differing evaluation cultures.

In addition, open and responsible research practices are still moving mainstream, and at different rates in differing countries, regions and types of institutions. Factors like levels of resources mean local contexts will have different levels of preparedness to adopt open and reproducible research practices. Since, we assert, it would be unfair for an institution to expect these practices before being able to adequately support their implementation (through training, services and infrastructure), it is essential that reforms are built upon adequate institutional foundations for performing open and reproducible research (Ross-Hellauer et al. 2022). In addition, particular open and responsible practices are of different relevance across disciplines, and hence reform must respect disciplinary cultures. Indeed, two recent surveys of research institutions by the European University Association highlights several barriers to change in research assessment from the institutional point-of-view (Morais and Borrell-Damian 2018; Saenen et al. 2019). Primary amongst these is the sheer complexity of the issue, which (as already stated) must account for differences in disciplines and career-stage, but also the various levels at which rewards and incentives can be structured, such as the level of research groups, departments, faculties, institutions, as well as (cross-)national actors like governments and research funders. Other factors identified by the EUA survey include lack of capacity, need to align policies with national or international agendas, resistance to reform from researchers or management, worries about increased costs, and lack of evidence on benefits (Morais and Borrell-Damian 2018; Saenen et al. 2019).

6 CONCLUSION

Current assessment of research and researchers forms a major barrier to uptake of open and reproducible research, retaining too much focus on inappropriate indicators, productivity as determined by quantity, and individual achievements rather than collaborative open research practices, and the socio-economic impact of research. In this paper we have quantitatively demonstrated that across countries, inclusion of such criteria remains rare. Although outreach to stakeholders beyond academia (public, industry, policy-makers) is somewhat better represented, practices related to Open Science are certainly not. Our sample is unique in addressing many countries across disciplines, and as such demonstrates that although general trends of over-quantification and undervaluing of open and responsible research can be observed, important differences between countries can be seen. In seeking reform, care must be taken to respect the historical and contextual reasons for such divergences.

6.1 Limitations and future work

The collection of policies was conducted under time constraints and in two rounds (one year apart). In addition, obtaining policies was difficult as they were often internal documents. Hence, there may have been some time-lag in that policies we examined were not the most recent versions.

The search protocol for the construction of the sample implied looking for the general RPT policy for the whole institution. This left out departmental policies and other policies such as Ethics, Diversity, and Open Access. In the Portuguese case, most Universities have a separate Open Access policy that in some cases is tied with promotion criteria, and therefore were not

included in our analysis. The focus on general policies might also explain the overall low rate of policies found for the USA.

All analyses presented in this paper are of exploratory nature. Our paper strives to explore the landscape of RPT policies to gather initial evidence on the prevalence of specific concepts. Future studies that focus on particular aspects and analyse them with a highly targeted approach (with pre-registered hypotheses of particular relationships) would be a meaningful extension of our work.

This work only identifies the prevalence of concepts in documents. It does not further analyse the particular contexts of use. Future qualitative work would provide this. In addition, this work does not consider how these policies were actually put into practice, and future survey or interview work would help in understanding these broader contexts. As an example, many policies mention that candidates should submit a full CV, including a complete list of publications. We did not consider these to be instances that quantify research output (indicator “Number of publications”). However, it is reasonable to assume that all requested materials will be considered to some extent, and that longer publication lists might be helpful.

A further factor to consider is, of course, the degree to which policies actually guide practice in review, promotion and tenure evaluations. As one of our anonymous reviewers astutely notes, some studies (e.g., Langfeldt 2001) indicate that reviewers are often highly selective in adhering to such criteria. If, for example, the removal of explicit reference to Journal Impact Factors is not

also followed by a cultural change whereby assessors are educated to the reasons they are not a good indicator of individual performance, then such factors may continue to play a role, even unofficially. Future work may build upon existing work (e.g., Hammarfelt 2017; Hammarfelt and Rushforth 2017) to further explore how criteria are implemented and weighted across disciplines.

Future work could also try to collect a larger number of RPT policies from a broader variety of geographical areas that are not included in our research. A greater sample would enable enhanced understanding of local contexts. Indeed, randomised sampling of a much broader range of countries might enable stronger claims about the state of criteria in use for RPT globally.

ACKNOWLEDGEMENTS

The authors gratefully thank the following for their contribution to the paper: Helene Brinken and Anja Rainer for assistance with data collection, and Bikash Gyawali and David Pride for assistance with data investigation. In addition, we kindly thank our two anonymous QSS reviewers and handling editor, Ludo Waltman, for their challenging and critical comments.

AUTHOR CONTRIBUTIONS

Nancy Pontika: Conceptualization, Data Curation, Investigation, Methodology, Project administration, Writing original draft, Writing - review & editing. Thomas Klebel: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. Antonia Correia: Data curation, Investigation, Writing – review & editing. Hannah Metzler: Conceptualization, Data Curation, Investigation, Methodology, Project administration, Writing - review & editing. Petr Knoth: Conceptualisation, Funding acquisition, Methodology, Project

administration, Supervision, Writing - review & editing. Tony Ross-Hellauer: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing.

COMPETING INTERESTS

The authors declare no competing interests.

FUNDING INFORMATION

This work was supported by the project "ON-MERRIT," funded by the European Commission under the Horizon 2020 program (grant number 824612).

DATA AVAILABILITY

All supporting data and required code are available in Zenodo (Pontika et al 2022b).

BIBLIOGRAPHY

- Adler, R., Ewing, J., & Taylor, P. (2009). "Citation Statistics." *Statistical Science* 24 (1): 1–14. <https://doi.org/10.1214/09-STS285>.
- Alperin, J. P., Muñoz Nieves, C., Schimanski, L., McKiernan, E. C., & Niles, M. T. (2019). "Terms and Concepts Found in Tenure and Promotion Guidelines from the US and Canada." <https://doi.org/10.7910/DVN/VY4TJE>.
- Alperin, J. P., Schimanski, L. A., La, M., Niles, M. T., & McKiernan, E. C. (2020). "The Value of Data and Other Non-Traditional Scholarly Outputs in Academic Review, Promotion, and Tenure in Canada and the United States." *Open Handbook of Linguistic Data Management*. <https://hcommons.org/deposits/item/hc:32909/>.
- Blasius, J., & Greenacre, M. (2006). "Correspondence Analysis and Related Methods in Practice." In *Multiple Correspondence Analysis and Related Methods*, edited by Michael J. Greenacre and Jörg Blasius, 4–40. Boca Raton: Chapman & Hall/CRC.
- Brechelmacher, A., Park, E., Ates, G., & Campbell, D. F. J. (2015). "The Rocky Road to Tenure – Career Paths in Academia." In *Academic Work and Careers in Europe: Trends, Challenges, Perspectives*, edited by Tatiana Fumasoli, Gaële Goastellec,

- and Barbara M. Kehm, 13–40. *The Changing Academy – The Changing Academic Profession in International Comparative Perspective*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-10720-2_2.
- Brown, J. D. 2014. "Citation Searching for Tenure and Promotion: An Overview of Issues and Tools." *Reference Services Review* 42 (1): 70–89. <https://doi.org/10.1108/RSR-05-2013-0023>.
- Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). "The Citation Advantage of Linking Publications to Research Data." *PLOS ONE* 15 (4): e0230416. <https://doi.org/10.1371/journal.pone.0230416>.
- Coonin, B., and Younce, L. (2009). "Publishing in Open Access Journals in The Social Sciences and Humanities: Who's Doing It and Why," 10.
- Costas, R., & Bordons, M. (2007). "The H-Index: Advantages, Limitations and Its Relation with Other Bibliometric Indicators at the Micro Level." *Journal of Informetrics* 1 (3): 193–203. <https://doi.org/10.1016/j.joi.2007.02.001>.
- Curry, S.. (2018). "Let's Move beyond the Rhetoric: It's Time to Change How We Judge Research." *Nature* 554 (7691): 147–147. <https://doi.org/10.1038/d41586-018-01642-w>.
- European Commission. Directorate General for Research and Innovation. (2017). *Evaluation of Research Careers Fully Acknowledging Open Science Practices: Rewards, Incentives and/or Recognition for Researchers Practicing Open Science*. LU: Publications Office. <https://data.europa.eu/doi/10.2777/75255>.
- Fleck, C. (2013). "Impact Factor Fetishism." *European Journal of Sociology* 2: 327–56. <https://doi.org/10.1017/S0003975613000167>.
- Gadd, E. (2021). *Mis-Measuring Our Universities: Why Global University Rankings Don't Add Up*. *Frontiers in Research Metrics and Analytics*, 6. <https://www.frontiersin.org/articles/10.3389/frma.2021.680023>
- Gardner, S. K., & Veliz, D. (2014). "Evinced the Ratchet: A Thematic Analysis of the Promotion and Tenure Guidelines at a Striving University." *The Review of Higher Education* 38 (1): 105–32. <https://doi.org/10.1353/rhe.2014.0045>.
- Greenacre, M., & Nenadic, O. (2018). *Ca: Simple, Multiple and Joint Correspondence Analysis* (version 0.71.1). <https://CRAN.R-project.org/package=ca>.
- Hammarfelt, B. (2017). "Recognition and reward in the academy: Valuing publication oeuvres in biomedicine, economics and history." *Aslib Journal of Information Management*, 69(5), 607–623. <https://doi.org/10.1108/AJIM-01-2017-0006>
- Hammarfelt, B. and Rushforth, A. D. (2017). "Indicators as judgment devices: An empirical study of citizen bibliometrics in research evaluation." *Research Evaluation*, 26(3), 169–180, <https://doi.org/10.1093/reseval/rvx018>
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). "Bibliometrics: The Leiden Manifesto for Research Metrics." *Nature News* 520 (7548): 429. <https://doi.org/10.1038/520429a>.

- Higginson, A. D., & Munafò, M. R. (2016). "Current Incentives for Scientists Lead to Underpowered Studies with Erroneous Conclusions." *PLOS Biology* 14 (11): e2000995. <https://doi.org/10.1371/journal.pbio.2000995>.
- Ioannidis, J. P. A. (2005). "Why Most Published Research Findings Are False." *PLoS Medicine* 2 (8): e124. <https://doi.org/10.1371/journal.pmed.0020124>.
- King, D., Acord, S. K., & Earl-Novell, S. (2010). "Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines." UC Berkeley: Center for Studies in Higher Education. <https://escholarship.org/uc/item/15x7385g>
- Landau, W. M.. (2018). "The Drake R Package: A Pipeline Toolkit for Reproducibility and High-Performance Computing." *The Journal of Open Source Software* 3 (21): 550. <https://doi.org/10.21105/joss.00550>.
- Langfeldt, L. (2001). The decision-making constraints and processes of grant peer review, and their effects on the review outcome. *Social Studies of Science*, 31(6). <https://doi.org/10.1177/030631201031006002>
- Malsch, B. 2015. "Journal Ranking Effects on Junior Academics: Identity Fragmentation and Politicization," *Critical Perspectives on Accounting* 26. <https://doi.org/10.1016/j.cpa.2014.02.006>.
- McGill, M. M., & Settle, A. (2011). "Computing Faculty Tenure and Promotion Requirements at USA and Canadian Post-Secondary Institutions." In *Proceedings of the 2011 Conference on Information Technology Education - SIGITE '11*, 133. West Point, New York, USA: ACM Press. <https://doi.org/10.1145/2047594.2047632>.
- McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., McDougall, D. et al. (2016). "How Open Science Helps Researchers Succeed." Edited by Peter Rodgers. *ELife* 5 (July): e16800. <https://doi.org/10.7554/eLife.16800>.
- McKiernan, E. C., Schimanski, L. A., Muñoz Nieves, C., Matthias, L., Niles, M. T., & Alperin, J. P. (2019). "Use of the Journal Impact Factor in Academic Review, Promotion, and Tenure Evaluations." Edited by Emma Pewsey, Peter Rodgers, and Björn Brembs. *ELife* 8 (July): e47338. <https://doi.org/10.7554/eLife.47338>.
- Migheli, M., & Ramello, G. B. (2014). "Open Access Journals and Academics' Behavior." *Economic Inquiry* 52 (4): 1250–66. <https://doi.org/10.1111/ecin.12131>.
- Morais, R., & Borrell-Damian, L. (2018). "Open Access in European Universities: Results from the 2016/2017 EUA Institutional Survey." 2018. <https://eua.eu/resources/publications/324:open-access-in-european-universities-results-from-the-2016-2017-eua-institutional-survey.html>.
- MoRRI. (2018). "The Evolution of Responsible Research and Innovation in Europe: The MoRRI Indicators Report D4.3." Technopolis-Group. 2018. <https://www.technopolis-group.com/report/the-evolution-of-responsible-research-and-innovation-in-europe-the-morri-indicators-report-d4-3/>.

- Nenadic, O., & Greenacre, M. (2007). “Correspondence Analysis in R , with Two- and Three-Dimensional Graphics: The **ca** Package.” *Journal of Statistical Software* 20 (3). <https://doi.org/10.18637/jss.v020.i03>.
- Niles, M. T., Schimanski, L. A., McKiernan, E. C., & Alperin, J. P. (2020). “Why We Publish Where We Do: Faculty Publishing Values and Their Relationship to Review, Promotion and Tenure Expectations.” *PLOS ONE* 15 (3): e0228914. <https://doi.org/10.1371/journal.pone.0228914>.
- Owen, R., Macnaghten, P., & Stilgoe, J. (2012). “Responsible Research and Innovation: From Science in Society to Science for Society, with Society.” *Science and Public Policy* 39 (6): 751–60.
- Peekhaus, W., & Proferes, N. (2015). “How Library and Information Science Faculty Perceive and Engage with Open Access.” *Journal of Information Science* 41 (5): 640–61. <https://doi.org/10.1177/0165551515587855>.
- Pinto, A. L., Matias, M., & Moreira González, J. A. (2016). “Produção da Ciência da Informação na Web of Science entre 1994 e 2013 e a lista Qualis/Capes da Área.” *Ibersid: revista de sistemas de información y documentación* 10 (1): 51–61. <https://doi.org/10.54886/ibersid.v10i1.4252>.
- Pontika, N., Knoth, P., Cancellieri, M., & Pearce, S. (2015). “Fostering Open Science to Research Using a Taxonomy and an ELearning Portal.” In *Proceedings of the 15th International Conference on Knowledge Technologies and Data-Driven Business*, 1–8. Graz Austria: ACM. <https://doi.org/10.1145/2809563.2809571>.
- Pontika, N., Klebel, T., Pride, D., Knoth, P., Reichmann, S., Metzler, H., Correia, A., Brinken, H., & Ross-Hellauer, T. (2021). ON-MERRIT D6.1 Investigating Institutional Structures of Reward & Recognition in Open Science & RRI (1.0). Zenodo. <https://doi.org/10.5281/zenodo.5552197>
- Pontika N, Gyawali B, Corriea A. et al. (2022a). How do career promotion policies affect research publications and open access? [version 1; peer review: awaiting peer review]. *Open Res Europe*2:99. <https://doi.org/10.12688/openreseurope.14921.1>
- Pontika, N., Klebel, T., Correia, A., Metzler, H., Knoth, P., & Ross-Hellauer, T. (2022b). Data for “Indicators of research quality, quantity, openness and responsibility in institutional promotion, review and tenure policies across seven countries”. Zenodo. <https://doi.org/10.5281/zenodo.6323765>
- R Core Team. (2021). *R: A Language and Environment for Statistical Computing* (version 4.1.2). Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Research and Innovation. (2022). “Process towards an Agreement on Reforming Research Assessment.” Text. European Commission - European Commission. 2022. https://ec.europa.eu/info/news/process-towards-agreement-reforming-research-assessment-2022-jan-18_en.

- Rice, D. B., Raffoul, H. Ioannidis, J. P. A. & Moher, D. (2020). "Academic Criteria for Promotion and Tenure in Biomedical Sciences Faculties: Cross Sectional Analysis of International Sample of Universities." *BMJ*, June, m2081. <https://doi.org/10.1136/bmj.m2081>.
- Rodriguez, J. E. (2014). "Awareness and Attitudes about Open Access Publishing: A Glance at Generational Differences." *The Journal of Academic Librarianship* 40 (6): 604–10. <https://doi.org/10.1016/j.acalib.2014.07.013>.
- Ross-Hellauer, T., Reichmann, S., Cole, N. L., Fessl, A., Klebel, T. & Pontika, N. (2022). "Dynamics of Cumulative Advantage and Threats to Equity in Open Science: A Scoping Review." *Royal Society Open Science* 9 (1): 211032. <https://doi.org/10.1098/rsos.211032>.
- Saenen, B., Morais, R., Gaillard, V. & Borrell-Damián, L. (2019). "Research Assessment in the Transition to Open Science," 48.
- Schomberg, R. v.. 2019. "Why Responsible Innovation." In *The International Handbook on Responsible Innovation. A Global Resource*. Cheltenham: Edward Elgar Publishing.
- Seipel, M. M. O. (2003). "Assessing Publication for Tenure." *Journal of Social Work Education* 39 (1): 79–88. <https://doi.org/10.1080/10437797.2003.10779120>.
- Sutton, E. (2020). "The Increasing Significance of Impact within the Research Excellence Framework (REF)." *Radiography* 26 (October): S17–19. <https://doi.org/10.1016/j.radi.2020.02.004>.
- Vinyard, M. & Colvin, J. B. (2018). "How Research Becomes Impact: Librarians Helping Faculty Use Scholarly Metrics to Select Journals." *College & Undergraduate Libraries* 25 (2): 187–204. <https://doi.org/10.1080/10691316.2018.1464995>.
- Walker, R. L., Sykes, L., Hemmelgarn, B. R. & Quan, H.. (2010). "Authors' Opinions on Publication in Relation to Annual Performance Assessment." *BMC Medical Education* 10 (1): 21. <https://doi.org/10.1186/1472-6920-10-21>.
- Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E. C. M., Tijssen, R. J. W., van Eck, N. J., van Leeuwen, T. N., van Raan, A. F. J., Visser, M. S., & Wouters, P. (2012). The Leiden ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology*, 63(12), 2419–2432. <https://doi.org/10.1002/asi.22708>
- Wang, J. (2014). "Unpacking the Matthew Effect in Citations." *Journal of Informetrics* 8 (2): 329–39. <https://doi.org/10.1016/j.joi.2014.01.006>.
- Wickham, H.. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Second edition. Use R! Cham: Springer.
- Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino McGowan, L., François, R., Grolemond, G. et al. (2019). "Welcome to the Tidyverse." *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.

- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S. Jones, E. et al. (2015). “The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management.” <https://doi.org/10.13140/RG.2.1.4929.1363>.
- Woolston, C. 2021. “Impact Factor Abandoned by Dutch University in Hiring and Promotion Decisions.” *Nature*, June. <https://doi.org/10.1038/d41586-021-01759-5>.
- Youn, T. I. K. & Price, T. M. (2009). “Learning from the Experience of Others: The Evolution of Faculty Tenure and Promotion Rules in Comprehensive Institutions.” *The Journal of Higher Education* 80 (2): 204–37. <https://doi.org/10.1353/jhe.0.0041>.