

Changing research on research evaluation: A critical literature review to revisit the agenda

Duncan A. Thomas^{1,2}, Maria Nedeva^{3,4,*}, Mayra M. Tirado³ and Merle Jacob⁴

¹Department of Political Science, Danish Centre for Studies in Research and Research Policy, Aarhus University, Bartholins Allé 8, 8000 Aarhus C, Denmark, ²Nordic Institute for Studies in Innovation, Research and Education (NIFU), PO Box 2815, Tøyen, N-0608 Oslo, Norway, ³Manchester Institute of Innovation Research, Alliance Manchester Business School, University of Manchester, Booth Street West, Manchester M15 6PB, UK and ⁴Department of Business Administration, Research Policy Group, Lund University School of Economics and Management, Lund University, PO Box 7080, 220 07 Lund, Sweden

*Corresponding author. Email: maria.nedeva@manchester.ac.uk.

Abstract

The current range and volume of research evaluation-related literature is extensive and incorporates scholarly and policy/practice-related perspectives. This reflects academic and practical interest over many decades and trails the changing funding and reputational modalities for universities, namely increased selectivity applied to institutional research funding streams and the perceived importance of university rankings and other reputational devices. To make sense of this highly diverse body of literature, we undertake a critical review of over 350 works constituting, in our view, the 'state-of-the-art' on institutional performance-based research evaluation arrangements (PREAs). We focus on PREAs because they are becoming the predominant means worldwide to allocate research funds and accrue reputation for universities. We highlight the themes addressed in the literature and offer critical commentary on the balance of scholarly and policy/practice-related orientations. We then reflect on five limitations to the state-of-the-art and propose a new agenda, and a change of perspective, to progress this area of research in future studies.

Key words: performance-based research evaluation arrangements; research evaluation; performance-based research evaluation; effects of research evaluation.

1. Introduction

In this article, we undertake a critical review of over 350 relevant publications that together constitute, in our view, a diverse and wide-ranging literature 'state-of-the-art' on the performance-based research evaluation arrangements (PREAs) of universities and other public research organizations. These arrangements¹ address systematic evaluation exercises aiming to introduce resource and reputational policy incentives aligned with dominant notions of research quality (Langfeldt et al. 2019). We believe our analysis is necessary to: (1) highlight major themes addressed by literature; (2) provide a critical commentary on the balance of scholarly and policy/practice-related orientations in this literature; (3) identify limitations in this state-of-the-art; and finally (4) propose a novel research agenda to overcome these limitations.

Evaluations of policy and funding arrangements to support public research have been undertaken and studied for many decades. However, the number of studies on the details and effects of specific research evaluation arrangements globally increased considerably during the 1990s. This growing interest trails changing funding modalities for universities and public research organizations, with a rise of competitive, project grant funding, increased selectivity applied to institutional research funding streams (Paradeise and Thoenig 2015), and the perceived importance of global rankings. Once pioneering research evaluation arrangements to allocate institutional funding, like Excellence for Research in Australia (ERA) and the UK Research Excellence Framework (REF), have also become established and seemingly intrusive enough to spur academic and policy concerns. This class of arrangements is becoming the

predominant evaluative means to allocate public research funds and/or garner global reputation. It is, therefore, our central focus.

These PREAs have been discussed in an increasingly large body of both academic and grey literature sources, addressed via both scholarly and more policy/practice-related orientations. The scope of this literature varies widely. There are small-scale studies on peer judgement and dynamics of peer review panels operating inside broader national PREAs. There is also research on wider effects for behaviours and strategies of actors, organizations and institutions in national policy, and funding ‘research spaces’, for example, for universities, funding agencies, and researcher career trajectories (see Nedeva 2013; see also Smith, Ward and House 2011; Waitere et al. 2011; Lee, Pham and Gu 2013; Aagaard, Bloch and Schneider 2015; Reale et al. 2018; Whitley, Glaser and Laudel 2018; Lind 2019).

Research attention has been paid to increasing selectivity, and increasing use of performance-based allocation approaches in institutional research funding in countries like Australia, the Netherlands, Sweden, and the UK (Organisation for Economic Co-operation and Development (OECD) 2009; Auranen and Nieminen 2010; Otley 2010; Wang and Hicks 2012; Tahar and Boutellier 2013; Leisyte and Westerheijden 2014; De Boer et al. 2015; Greenhalgh and Fahy 2015; Jonkers and Zacharewicz 2015; Arocena, Göransson and Sutz 2018; Canibano et al. 2018; Jonkers and Sachwald 2018; Woelert and McKenzie 2018). PREAs have also become of central importance in terms of research/epistemic governance. There is a perceived transition away from determination of research goals and orientations endogenously within universities and knowledge communities towards greater authority and influence from more strategic and managerial policy and university actors designing, deploying, or reacting to outcomes of PREAs (Whitley and Gläser 2007; Langfeldt et al 2019).

Our critical review and analysis of this literature aims to identify thematic coverage, highlight limitations, and propose a new research agenda we believe is needed to move studies forward in this area. There have been previous surveys of research evaluation-related literature, for example, cross-sectional surveys and thematic reviews of evaluation practices and indicators (see De Rijcke et al. 2016). There have also been comprehensive studies correlating specific characteristics of differing national research evaluation arrangements to apparent national science system performance or excellence in international context (see Sandström and Van den Besselaar 2018; also Jonkers and Sachwald 2018). Whilst remaining within the confines of a critical review approach, our intent here is different and somewhat closer to meta-research motivations (c.f. Ioannidis 2018). We aim to analyse the themes, orientations, and limitations of research evaluation *research* itself, a review approach we believe has been overlooked in literature in this area to date.

In doing so we grapple with a messy reality. PREAs are dynamic, often politicized and are not ‘scientific’, static, standardized, or universal. They operate across multiple spatial levels and time horizons, use differing methods, involve varying degrees of transparency and costs, and are conducted by different kinds of organizations for various purposes (Galleron et al. 2017). They can be understood as socially constructed systems, their legitimacy and effectiveness can be disputed, and they blend multifaceted contextual, political, managerial, economic, and reputational elements (Bianco, Gras and Sutz 2016). We believe our critical review and analysis must therefore be *purposive* rather than trying to encompass all possible research on this vast topic.

To structure our article, first, we define our understanding of PREAs and use it to guide our approach. We describe our purpose in collecting and coding a bespoke dataset of 354 pieces of literature that we believe constitutes the most relevant ‘state-of-the-art’ on PREAs. Second, we present an analysis of five research themes we derive through inductive clustering of this state-of-the-art and provide critical commentary on the major arguments in this body of research. Third, we discuss five limitations to this PREA-related literature and suggest a novel research agenda to address them.

2. Approach

We understand PREAs as including ‘organized sets of procedures for assessing the merits of research undertaken in publicly funded organizations that are implemented on a regular basis, usually by state or state-delegated agencies’ (Whitley and Gläser 2007: 6). PREAs operate at multiple levels, as an ‘ensemble of practices and institutional arrangements in a country’ and/or locally in a university organization, mediating ‘between scientific quality controls and research policies’ (Cruz-Castro and Sanz-Menéndez 2007: 205). They are part of the ‘organizational governance’ of universities, in directing ‘strategy, funding’ and operations, and are a potential source of tensions (Luo, Ordóñez-Matamoros and Kuhlmann 2019: 1). They are also frequently ‘intended to change science by improving its quality’ and possibly even altering research ‘content’ (Gläser 2007: 245). They can be ‘weak’ and aim primarily at ‘information-gathering’ for benchmarking of research, researchers, and research organizations—or else ‘strong’ in performance-based ‘national systems of research output evaluation’ and be used as a basis ‘to distribute research funding to universities’ (Hicks 2012: 260).

To guide our critical review and analysis of the literature, we capture the most salient of these aspects by defining PREAs here as the *institutionalized, or semi-institutionalized, practices and procedures aiming to assess the merit of the research output, research environment, and research engagement of research organizations with a view to incentivizing desired change or continued performance*. PREAs may be conducted at different levels of social aggregation—for example, national research system, organization, etc.—and affect resource allocation and reputations.

Not to conflate our definition of PREAs with other possible forms of evaluation, we draw upon an understanding of science dynamics as involving research fields and research spaces (Nedeva 2013). We thus distinguish between PREAs and two other commonly addressed types of research evaluation. Our critical review includes only literature on PREAs located in the research space (see Figure 1). We thereby exclude literature addressing research evaluation types performed by research organizations and research field-related knowledge claim assessment.

This PREA definition directed our focus to the field of science, technology, and innovation policy (STIP) studies, and mapped onto central and peripheral journals in this area. Our approach followed that of a *critical narrative review*; we wished to identify key contributions around our specified topic but not necessarily to address all evaluation-related material ever produced (c.f. Demiris, Oliver and Washington 2019). Our definition directed us to core STIP-related journals (e.g. *Research Evaluation*, *Research Policy*, *Science and Public Policy*, *Scientometrics*, and *Minerva*) and selected peripheral ones.²

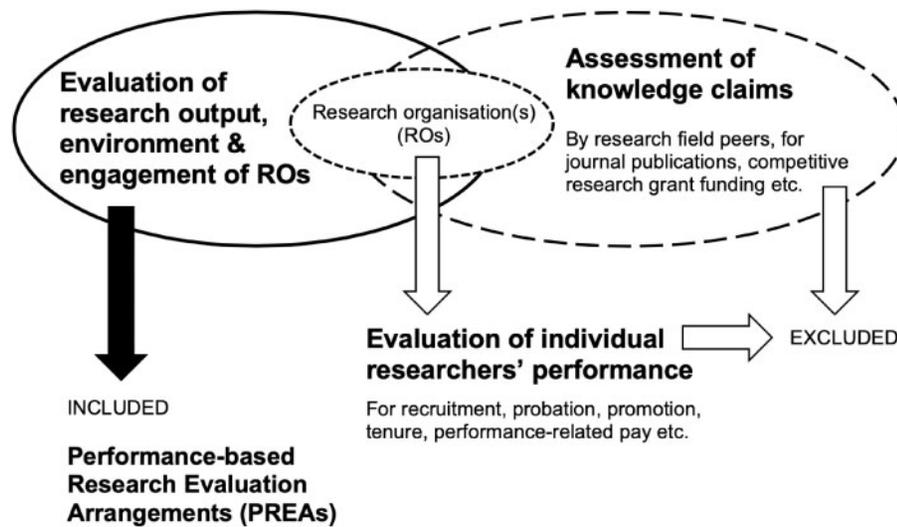


Figure 1. Schematic of our literature critical review search strategy to include PREAs and exclude other evaluation types (adapted from Nedeva 2013).

We used the keywords ‘research evaluation’, ‘institutional’, and ‘university(ies)’ in searches of (1) Web of Science, (2) Scopus, and (3) Google Scholar. This resulted in 675 hits from numerous journals, books, and non-academic sources. We reviewed titles and abstracts at this stage to screen for duplicates and, guided by our PREA definition, ensured materials primarily addressed research space-related research evaluation. This was done using (1) our knowledge as active scholars in fields of research evaluation and research policy for several decades (c.f. Adler and Adler 1987); (2) our knowledge of research consultancies and their key reports; and (3) invited expert advice by email, telephone, and face-to-face from a small number of international research policy/evaluation academic and consultant colleagues (this latter element introduced an element of *consensus* narrative review; c.f. Wilczynski 2017). Our final set thus also included grey literature from consultancies and funders like Technopolis, PA Consulting, the European Commission, the former Higher Education Funding Council for England (HEFCE), and select others.³

This critical narrative review process with an element of consensus review led to our final set of 354 full-text materials, including academic articles, books, funder and policy reports that we then inductively coded and analysed. The earliest piece of literature that we retrieved was published in 1968. For convenience, we set 2018 as a cut-off publication year. Just over 85% of the literature we included and reviewed in this bespoke dataset was published between 2000 and 2015—reflecting increased attention as funding modalities and evaluation arrangements have been recently changing. A total of 179 items were primarily qualitative, 103 were quantitative, and 72 were mixed methods based. The literature in the dataset addressed PREAs related to 37 countries and territories, trans-national arrangements, and international surveys of these arrangements (e.g. by the European Union [EU] and the Organisation for Economic Co-operation and Development [OECD]).⁴

Following this highly selective, expert-informed, critical and consensus narrative review approach we cannot claim to have produced a comprehensive collection of *all* materials ever published on ‘research evaluation’-related topics. However, we believe we captured enough breadth and depth of the ‘state-of-the-art’ on PREA-related

topics to satisfy our purposive analysis, to highlight key limitations, and to underpin our proposition of a novel research agenda.

For every piece of literature in the dataset we manually read abstracts and full texts. From this reading, we wrote synopses summarizing the approach, coverage, findings, and conclusions of each piece of literature. We then analysed our database of synopses to produce an inductive clustering of all the literature into *five major themes*, shown in Table 1. All literature was assigned to a single major theme based upon primary message. This was based on our subjective reading of the literature content, what proportion of it addressed a given theme, and the prominence afforded that theme in the literature.⁵

Our first inductive clustering theme, *accounts of local PREAs* was where we assigned literature whose primary content provided ‘thick descriptions’. This included case studies of PREAs specific to a national research system (e.g. ERA or REF), a trans-national regional bloc (e.g. EU-level arrangements), for a sub-national region, for a specific organization (e.g. university), or for a sector or grouping of organizations (e.g. medical research in universities and research institutes). Our second theme was where we clustered *comparative studies of PREAs*, for instance, those comparing specific sets of countries or specific research fields. Our third theme captured literature providing *discussions of rationales for (performance-based) research evaluation*, for example, discussing the policy impetus for performance-based criteria and how they related to pursuit of excellence aims, efficiency, and other concerns. The fourth theme clustered *appraisals of (performance-based) research evaluation methodologies*; for example, debates around the relative merits of bibliometrics, altmetrics, and other indicators vis-à-vis peer review practices—essentially the detailed methods and machinery, technical parameters, and logistics of the design and deployment of PREAs. Our fifth and final theme clustered literature attempting *studies of effects on the science system*, for example, how PREAs interacted with science dynamics and researcher careers.

We found it helpful to characterize the literature further using limited additional coding: literature *type* (i.e. journal articles, books or book chapters, policy reports); literature *content*—primary research (e.g. interviews, surveys, bibliometrics, mathematical models

Table 1. Overview of the five major themes we produced to cluster our database, and further details of the 354 pieces of PREA-related literature

Theme	Description	Number of contributions	Years published
1	Accounts of local PREAs	100 (28%)	1990–2018
2	Comparative studies of PREAs	40 (11%)	1994–2016
3	Discussions of rationales for (performance-based) research evaluation	18 (5%)	1994–2018
4	Appraisals of (performance-based) research evaluation methodologies	103 (29%)	1971–2018
5	Studies of effects on the science system	93 (26%)	1968–2018

and simulations, mixed methods) or secondary (e.g. desk-based literature reviews and/or secondary sources); literature *methods* (quantitative, qualitative or mixed); literature *approach*, that is, thick descriptions of specific cases, critical analyses, and attempts at comparative analysis; and *object* of analysis, that is, organization level evaluations or sub-national, national, or trans-national levels.⁶ These further codes are shown in Table 2 and were included in our analytical approach.⁷

Inductively clustering these five themes and using our further coding we began our purposive analysis, where we posed five specific questions:

- What key themes have been addressed by this literature?
- What is the balance of research attention across all the themes?
- What are the analytical implications of the apparent balance between scholarly and policy/practice-orientations in this literature?
- What aspects have not been addressed?
- Given this state-of-the-art, what new research agenda might move PREA-related research forward?

3. Findings

We now present our analysis of the dataset of 354 pieces of literature. For each of the five themes, we provide a summary of key research arguments, brief critical commentary, and descriptive information using our further codes.⁸

3.1 Theme 1: Accounts of local PREAs

Theme 1 grouped literature we determined to be primarily focused on providing descriptive *accounts of local PREAs*. Altogether we assigned 100 pieces of literature to this theme. A total of 77 pieces described PREAs at *national* level, for example, national evaluations like those in Australia, the Netherlands, and the UK. Within this theme, we also placed literature primarily describing arrangements at *organizational* (six pieces of literature), *sub-national* (five pieces), and *trans-national* levels (11 pieces).⁹

National level PREAs were described for countries where these practices were already well established, like the UK (Barker 2007; see also Martin and Whitley 2010; Morris 2010) and Australia (Butler 2008; Donovan 2008). These arrangements were also described in other literature, to show them as apparent exemplars for development and implementation of new arrangements in countries or regions that had previously not used such practices (Fiala 2013; Ancaiani et al. 2015; see also European Centre for Strategic Management of Universities (ESMU) 2010; Geuna and Piolatto 2016). Some literature included not only primary descriptive content but also secondary messages, such as critical feedback to policymakers, and reflections on existing arrangements for possible policy

learning (Hare 2003; Adams and Gurney 2010; Elsevier 2013; Hughes, Kitson and Bullock 2013; Van Drooge et al. 2013; Higher Education Funding Council for England (HEFCE) 2014; Wouters et al. 2015; see also Henkel 1999; Auranen and Nieminen 2010; Broadbent 2010; Elsevier 2011; Spaapen and Van Drooge 2011).

Literature we grouped in Theme 1 had often been commissioned by national or international organizations responsible for evaluating research outputs, environments and engagements of higher education institutions or other research organizations. Nearly half the material in Theme 1 (46 pieces of literature) was policy reports describing national-level arrangements, then benchmarking them against each other to provide an international overview. These kinds of policy reports were commissioned and (presumably) funded by ministries of education in different countries, the OECD and say, the former HEFCE in the UK. We determined these bodies had funded these studies to enable policy learning about past experiences and/or arrangements used in other countries.

The bulk of Theme 1 literature we would call ‘highly descriptive’ (81 pieces of literature). We determined they used no explicit theoretical positions. A similar number used primarily qualitative and/or mixed methodologies (81 pieces). Ten pieces of Theme 1 literature had what we would consider more analytical approaches; 19 used quantitative methodologies, for example, Cattaneo, Meoli and Signori (2016) (see also Frølich 2008, 2011; Frølich, Schmidt and Rosa 2010; Wang and Hicks 2012; Frankel, Goddard and Ransow 2014; Hamann 2016); and 42 of the 100 pieces collected primary data. The others based their descriptive accounts on secondary research and sources.

The descriptions of PREAs across Theme 1 literature addressed the following: descriptions of national-level arrangements (broad and fine details); evaluation strategies (apparent purposes, economic and social rationales); funding mechanisms (i.e. whether and how much evaluation results were linked to funding streams); assessment methods and inclusion/exclusion criteria of what was assessed; how often assessment took place; what units were assessed (research themes, research organizations, etc.); and evaluation outcomes (e.g. apparent levels of research-related performance of organizations, regions or nations, based on indicators such as publication volumes, citations, number of patents, and/or university–industry links). Theme 1 literature primarily used case study research designs and detailed the—sometimes considerable—costs associated with (repeated) use of research evaluation. Some provided cost-benefit analyses of existing evaluation exercises (e.g. see Campbell and Boxall 2004; PA Consulting Group 2008; Technopolis 2009, 2010; see also Mahieu, Arnold and Kolarz 2013, 2014; Arnold et al. 2014; Mahieu and Arnold 2015). We classed these pieces of literature as largely ‘user-driven’. They seemed designed to answer research questions or address research interests of policymakers and evaluation practitioners.

Table 2. Description of our PREA-related literature database coding

Code	Value(s)	Analytical rationale
Theme	As per Table 1	Identify major thematic contributions on PREAs and their distribution
Type	Journal article; book/book chapter; policy report (i.e. grey literature)	Identify distribution of publication types
Content	Primary research; secondary research	Highlight any prevalence regarding data collection
Methods	Qualitative; quantitative; mixed	Highlight any prevalence of method
Approach	Thick description; critical analysis; comparative analysis	Highlight any prevalence of analytical approaches
Object	Organizational evaluation; sub-national evaluation; national evaluation; trans-national evaluation; other (theory, instruments, etc.)	Register coverage of objects of analysis

Turning a critical eye to Theme 1 literature, we found an absence of *frameworks* for theoretically or conceptually based study and analysis of PREAs. Theme 1 literature was primarily *descriptive*—both for the material published in academic journals and ‘grey literature’, user-driven, policy reports. This potentially presents a problem and may not be an ideal basis to support robust policy learning. This literature in our critical opinion *does not* provide analysis and comprehension of social mechanisms around PREAs. However, it clearly *does* provide a source of rich empirical material and cases that could later be revisited for analytical purposes.

3.2 Theme 2: Comparative studies of PREAs

We assigned 40 pieces of literature into our clustering Theme 2. These were *comparative studies of PREAs*, comparing, for example, arrangements for specific sets of countries, or for particular research fields. Some undertook broad comparisons of institutional and other evaluation arrangements (Geuna and Martin 2003; Orr 2004; Hicks 2010; Arnold and Mahieu 2015; see also Frølich 2008; Geuna and Piolatto 2016; Sandström and Van den Besselaar 2018). Others compared selective research funding arrangements, effects for behaviours like research collaboration (Johnston 1994), actions of research funding agencies (Lepori et al. 2009), consequences of evaluation for university funding (Franzoni, Scellato and Stephan 2011; see also Sörlin 2007), or PREA-related criteria for assessing research quality in different fields (Hug, Ochsner and Daniel 2013).

Literature here provided accounts of PREAs in multiple different settings and countries, but crucially with few attempts at *analytical* comparison. Hicks (2010), for instance, compared specific research evaluation *objectives* and *strategies* used by EU countries, Australia, South Africa, and some Asian countries—but did not compare wholesale the design, operation, and effects of these arrangements within a comprehensive framework. Reborra and Turri (2013; see also Geuna and Piolatto 2016) compared how research funding of universities evolved over time to incorporate *selectivity* and *evaluation* elements, specifically in the UK and Italy. Similarly, Geuna and Martin (2003) compared specific *methods* of evaluation used in 12 countries in Europe and the Asia-Pacific region.

Like Theme 1, the majority of Theme 2 literature we considered *user-driven* policy reports (26 pieces of literature or 65% of this theme was policy reports; 14 were academic publications, i.e. journal articles, a book, and a book chapter). Some Theme 2 literature also compared PREA-related practices across different countries to support policy learning (Iorwerth 2005; Grant 2010) or as guidance for policymakers wishing to implement and institutionalize PREAs in new settings (see e.g. Arnold and Mahieu 2015). Theme 2

literature was largely based on secondary research (in 29 pieces or 73% of Theme 2) and used qualitative or mixed research methods (88% of literature in Theme 2).

3.3 Theme 3: Discussions of rationales for (performance-based) research evaluation

Literature we clustered into Theme 3 primarily provided *discussions of rationales for research evaluation*, for example, the policy impetus and rationales for using performance-based evaluation criteria or how policy concerns and performance criteria like excellence and efficiency were interrelated. This Theme 3 was a very specific subset of the literature. It was our smallest cluster, at only 18 pieces.¹⁰ Some analytical frameworks were present in Theme 3 but no common or shared framework was used across different literature here.¹¹

A first key argument in the Theme 3 literature was that the introduction of PREAs requires that one also consider value-for-money and issues of research quality. Here, Theme 3 literature suggested policymakers’ rationales included values like promoting knowledge-based economies and strongly overlapped with efforts to use public research systems in different national settings to revive and/or restructure the orientation and/or performance of whole national economies (e.g. Rip and van der Meulen 1996; Bozeman and Sarewitz 2011; Sørensen, Bloch and Young 2015; Jonkers and Sachwald 2018; see also Elton 2000; Geuna and Martin 2003; Bence and Oppenheim 2005; Frølich, Schmidt and Rosa 2010; Martin and Whitley 2010; Mingers and White 2015; Woelert 2015).

A second key argument was that PREAs are evolving in parallel with rationales asserting that more competitive allocation of research funding improves research performance, for example, as judged by measures like publication productivity, and other indicators of apparent ‘excellence’. Theme 3 therefore seemed to include an emerging, critical research tradition moving close to addressing effects of competitive funding interventions as part of evolving PREAs. The interweaving of competitive funding and research evaluation was treated from research funders’ perspectives, at national research system level and in some cases at the level of researchers (Benner and Sandström 2000; Smith, Ward and House 2011; see also Sørensen, Bloch and Young 2015).

Theme 3 literature suggested consideration of PREAs has to account for public research funding becoming more fine-grained over time. Previous research funding regimes generally treated most if not all aspects of the research system like a ‘black box’. For instance, literature here described ‘first generation’ institutional research funding streams that did not address researchers, but simply took

universities in the aggregate. Later approaches channelled funding streams by disaggregating research systems into actors, institutions, etc. and attempted to leverage specific types of outcomes or promote specific behaviours. Theme 3 literature documented and critiqued this shift, noting the journey of the word ‘excellence’, in particular, from being an idiosyncratic, field-specific term used by researchers to recognize extraordinary scientific contributions, to it becoming an indicator used by policymakers and university managers to refer to *all* research as *sharing* some predefined set of characteristics (e.g. Laudel 2005; Smith, Ward and House 2011; see also Hicks 2012; Sousa and Brennan 2014). For example, Sørensen, Bloch and Young (2015) concluded that when ‘excellence’ was discussed in the context of PREAs it had now moved from being a marker of purely scientific performance to a broader basket of additional research performance-related criteria, for example, potential commercialization of research outputs, and indeed anything ‘commercializable’.

A third key argument in Theme 3 literature was a travel of global policy and economic competitiveness discourse into PREAs. The rise and diffusion of ideas (and ideology) around the global competition for knowledge, resource constraints, and resultant changing views of universities were chronicled, that is, a change from them being civic, public organizations to being more like corporations, and venues where performance must be audited. Theme 3 literature considered how conceptions of knowledge have shifted, and excellence has become a means within PREAs to reward ‘winners’ and punish ‘losers’. This was described as a new ‘strategic approach’ to research policy and resource allocation through these arrangements, suggesting policymakers and governments have moved closer, in theory if not yet in practice, to selecting and affecting the types and topics of research, research content (methodologies, equipment), and even *which* specific researchers they believe can deliver ‘excellence’ within a particular research system (Benner and Sandström 2000; Sörlin 2007; see also Hicks 2012; Watermeyer 2014, 2016).¹² An apparent merging was noted, of policymakers’ search for ‘excellence’ and use of evaluation as a tool to measure research system effectiveness, with guiding and directing socio-economic investment decisions.

All but one piece of literature in Theme 3 was published in academic journals. Theme 3 literature drew mainly on secondary data, used qualitative methods, and was the most analytical set, in our view. Literature here attempted to unpack varying, evolving rationales for PREAs, and to trace how they were now being seen as enablers of structural change, and as facilitating national systems that could compete more at an international level.

3.4 Theme 4: Appraisals of (performance-based) research evaluation methodologies

Nearly a third of all the pieces of literature in our database (103 pieces, 29% of the full dataset) addressed methods related to PREAs, for example, whether and which indicators were reliable measures or proxies to evaluate research performance, in terms of excellence and quality (Cozzens 1981; Donovan 2007; De Jong et al. 2011; Wunsch-Vincent 2012; Wilsdon et al. 2015; see also Aagaard 2015). These pieces we clustered in Theme 4. Literature here we judged as aiming to discover or design the ‘best’ methods for PREAs to assess subjective notions like research excellence and quality. Some favoured exclusive use of peer review or of bibliometrics. Others advocated mixed approaches say, combining peer review and bibliometrics techniques (Butler 2007; Abramo, D’Angelo and Di Costa 2008, 2011; Abramo and D’Angelo 2011; Franceschet

and Costantini 2011; Abramo, Cicero and D’Angelo 2013; Eyre-Walker and Stoletzki 2013).

Theme 4 literature was very useful in highlighting two current dilemmas around design and deployment of differing PREAs. First, materials here considered *which* approach should be used, that is, predominantly qualitative or quantitative? Some literature addressed whether qualitative peer review was the most appropriate and/or cost-effective instrument to use or whether use of bibliometrics and other kinds of quantitative indicators was preferable. Other literature advocated use of blended or mixed approaches. Bertocchi et al. (2015), for instance, suggested research performance be evaluated using bibliometrics as an initial input for subsequent peer review. Still others proposed bibliometrics be used at national or local level to manage and/or monitor research performance within an evaluation, before feeding into later large-scale, peer review-based judgements, that is, so-called ‘informed’ peer review (see Neufeld and von Ins 2011; Wilsdon et al. 2015).

A second dilemma in Theme 4 literature was *how* current methodologies might be modified for use by policymakers and/or university managers to *encourage*, or at least not *impede* sustainable research activity in specific fields (e.g. in social sciences and humanities, SSH) or to foster research with particular properties (e.g. breakthrough, frontier, long-term). For instance, in SSH ‘informed’ peer review was advocated to assess better the performance of research fields where publishing journal articles represent only *part* of research outputs activities (e.g. in political science, where books and policy engagement also occur, Donovan 2009). Other literature suggested the same approach be part of PREAs in fields where peer review was dominated by reviewers representing only specific sub-fields (e.g. all denominations of economists being evaluated only by neoclassical/mainstream economists; Lee and Harley 1998; Lee, Pham and Gu 2013). Theme 4 literature advocated or designed new field-specific, more ‘inclusive’ quantitative indicators (e.g. social media-related ‘altmetrics’) to account for societal effects, broader or ‘alternative’ research outputs, interactions, exchanges, and outcomes (Bozeman, Dietz and Gaughan 2001; Kaufmann and Kasztler 2009; Kenna and Berche 2011; Ochsner, Hug and Daniel 2012; Kwok 2013; Sastry and Bekhradnia 2014).

Theme 4 literature predominantly featured material published in academic journals (83 of the pieces or 80% of Theme 4), relied on secondary data (80 pieces) and used quantitative methodologies (58 pieces). The predominant object of analysis was PREAs at *national* level (in 63 pieces of literature).

3.5 Theme 5: Studies of effects on the science system

Our final Theme 5 covered *studies of effects on the science system* from PREAs. Here, we clustered 93 pieces of literature, addressing effects at multiple spatial levels (regional, national, trans-national) and analytical levels (system, organization, researcher, research topics and content). Some literature instead took a cross-cutting view across these levels. Effects of PREAs on *universities* specifically were a dominant focus. Other literature combined this with attention to a general shift away from institutional/block funding towards proportionally more of competitive, project-based research funding allocation. Few pieces of literature addressed effects of PREAs upon *additional* parts of the science system *beyond* universities, say, effects for global research fields or aggregate effects at global level of multiple differing arrangements operating in parallel at national and/or regional levels.

Some Theme 5 literature argued specific PREAs have generated effects at the ‘macro’ level of changing how science, universities, and scientists/researchers are perceived by society. The critical view was that strategic use by policymakers and university managers of particular arrangements—with perhaps disproportionate emphasis here upon the UK’s Research Assessment Exercises (RAEs) and REF—had significantly changed organizational conditions for, and authority relations around knowledge creation (Himanen et al. 2009; see also De Jong et al. 2011; Kallerud et al. 2011; Whitley, Glaser and Laudel 2018).

At ‘meso’ level, literature observed that publicly funded research universities had become vulnerable to, and at risk of, being transformed by what certain exogenous stakeholders (e.g. politicians, policymakers, research funding agencies, corporate actors) considered ‘best’ for them. They were portrayed as losing autonomy, scholarly leadership, and ability to generate new and/or critical academic ideas. Universities and their researchers were framed as forced to abandon Mertonian notions of autonomy, disciplinarity, and freedom (c.f. Merton 1968) and expected to adopt values and quality standards shaped by outside demands (Frølich, Schmidt and Rosa 2010; Harland et al. 2010; see also Luukkonen 1997; Van der Meulen 1998; Ferlie, Musselin and Andresani 2008). Universities were diagnosed as no longer doing what they were ‘best’ at, and as complying with exogenous quality and excellence standards imposed by PREAs—or forced to suffer consequences of reduced research revenue and/or national and global reputation in local and world rankings/league tables (Knowles and Burrows 2014; see also Elton 2000; Luukkonen and Thomas 2016).

Other effects on universities included university management practices described as moving away from traditional ‘academic’ values (Linkova 2014; Agyemang and Broadbent 2015), changed university hiring, probation, and promotion strategies, allied to university strategic objectives and management practices becoming strongly coupled to criteria derived from evaluation-related goals and targets (see also Henkel 1999). Universities were also framed as embracing competition rather than resisting it and using PREAs at ‘micro’ level, to develop and deploy incentives, and ever more granular research information systems, monitoring and auditing mechanisms, to foster, reward, or sanction particular kinds of research productivity by research groups and at individual researcher level (Nedeva et al. 2012).

Other reported ‘meso’ level effects were university management game-playing, particularly within ‘strong’ PREAs directly linked to resource allocation (Whitley, Glaser and Laudel 2018). Universities, their leaders, and managers were reported as developing and using deliberate strategies to incentivize and direct types of research, researchers, and external university-stakeholder relationships that painted them in the most favourable light within PREAs so as to maximize research funding capture (again, particularly relating to the UK’s RAEs/REF). This behaviour reportedly has led to: undesirable concentration of resources by funders and universities to support short-term ‘safe’ rather than long-term risky research; allocation of resources to meet lay stakeholder/proxy indicators of excellence irrespective of knowledge community/substantive judgments about research quality; favouring competition over collaboration, thus risking fragmentation of academic/professional collegiality and reciprocity within and across universities; and direct or indirect promotion of ‘salami slicing’ publication practices to reward publication of a greater quantity of perhaps less comprehensive research works rather than focus on fewer but potentially more

significant publications of ‘higher’ quality (Butler 2003; Leisyte and Westerheijden 2014; see also Abramo, D’Angelo and Di Costa 2011).

Further effects were reported to be: increased short-termism generally at universities; superficial attention to what in some quarters are seen as spurious markers of university reputation/excellence in national and global league tables for universities ‘playing the game’; erosion of creativity; reduced diversity of the research topics, methods and approaches researchers’ pursue; and strategy and management level distortions in resource allocations that undermine previous synergies between teaching and research (Whitley, Glaser and Laudel 2018; see also Paradeise and Thoening 2015). Some authors even felt ‘strong’ PREAs (i.e. coupled to funding allocation) and audit cultures ‘dehumanized’ researchers and harmed traditional, more liberal, long-standing purposes and roles of universities in wider society (Hare 2003; Harland et al. 2010; Olssen 2016; see also Geuna and Martin 2003; Martin and Whitley 2010).

Some Theme 5 literature addressed effects at the ‘micro’ level of researchers and their research work processes: apparent loss of academic work-life balance and freedom; downgrading of teaching relative to research/publications; loss of intellectual curiosity; and a debasing of the general character of academic scholarship (Court 1999; Roberts 2007; Linkova 2014; Vincent 2015). Reported centralization of authority towards organizational elites like university managers, using expanding research data systems and information sourced from national/external and local/internal PREAs, were considered avenues of (negative) control over research content (Gläser et al. 2010; see also Aagaard 2015). PREAs were also reported to increase administrative burdens for researchers and decrease research time and productivity (Martin 2016, 2011).

Other Theme 5 literature indicated a fundamental transformation *cutting across* macro/meso/micro levels that had reportedly changed: university (research) culture; the nature, remit, processes, and practices of universities’ objectives and goals; the relevance of university research; and research topic coverage and diversity. These effects were linked to changing university strategies to mobilize the outcomes of PREAs to improve positioning in university rankings (Martin 2011; Holmes 2015). Academia and knowledge were described as being reconceptualized as commodities, driven by economic efficiency and value-for-money concerns. A shift towards performativity was reported, with universities and academics assigned and/or adopting new purposes within these changing authority relations (Harland et al. 2010; Whitley 2011). These relations included policymakers, and university managers, administrators, and field elites in universities using their newfound authority to attempt to ‘steer’ science systems even at the expense of marginalizing input from academics and other voices. Some authors here sounded a ‘wake-up call’ for academics to resist supposedly harmful use of PREAs and fight to retain long-held values that give meaning to ‘the academy’ (Martin and Whitley 2010; Martin 2011, 2016; Waitere et al. 2011; see also Bence and Oppenheim 2005; Murphy and Sage 2014). Authors contended PREAs should prove *their* usefulness in improving research culture, financial sustainability, research capacity, and so on in universities—rather than that academics should bow and bend to fit better the parameters of these arrangements. Some authors here foreshadowed an ‘end’ to universities as places for reflection and creative thinking, extinguished by the utilitarian influence of PREAs—even those PREAs that advocate and incentivize seemingly more positive societal ‘impact’ from research

(Knowles and Burrows 2014; see also Claeys-Kulik and Estermann 2015).

Other Theme 5 literature reported changes to the global communication system of science. Academic journal editors were reported as developing strategies to inflate their own journal rankings and citation counts to pander to use of PREAs and thus to become more attractive to authors (Gibson, Anderson and Tressler 2014). Journal editors were criticized for apparently seeking fewer path-breaking, critical research ideas and methods to publish (that reportedly accumulate citations more slowly), instead favouring more immediately citable, fashionable topics and approaches that can quickly inflate journal impact factors. Some Theme 5 literature described academic editors, publishers, reviewers, universities, government, and funding agencies as *collectively* adapting here to PREAs (Macdonald and Kam 2010; Watermeyer 2016).

We make two main critical points about this Theme 5 literature. First, *little is known about causal relationships* between PREAs and *many* if not *all* of these reported changes and apparent effects (see also Gläser 2019). This holds true for micro-level changes in research topic selection and researchers' pursuit of research programmes/lines and for other levels (Waitere et al. 2011; De Rijcke et al. 2016; Hammarfelt and de Rijcke 2015; see also Laudel 2005; Whitley and Gläser 2007). There are inherent methodological difficulties to measure and attribute PREA-related change here within and across heavily mediated, multi-level, multi-actor, regional, national, and trans-national research funding and policy 'spaces' and global 'research fields' (Nedeva et al. 2012; Whitley, Glaser and Laudel 2018).

Second, this literature may be *biased by over-representation of both scholarly and more personal accounts/normative responses* to the UK RAEs/REF. The UK's primary PREA is globally influential, but we must remember it is not necessarily 'best practice', has not travelled to many other regions of the world, and analytically the UK is an outlier or 'unique' (Sivertsen 2017). Reported effects there cannot be taken to be representative of effects of differing arrangements in other contexts (this criticism of course also ties in with the lack of comparative analytical frameworks across the literature state-of-the-art). There are few attempts to distinguish analytically the RAEs/REF from other PREAs or to make theory-based assumptions and arguments to link causally particular arrangements to specific effects.

In overview, most Theme 5 literature was published in academic journals (82 pieces of literature or 88% of this theme). Many arguments were built on either primary (43 pieces) or secondary data (50 pieces) and used qualitative approaches (in 61 pieces of literature). We considered most Theme 5 literature to be predominantly *analytical* in approach (54 pieces).

3.6 Cross-cutting issues

Looking across all five clustering themes most literature seemed to share the view that, whatever the specific arrangements, PREAs are 'here to stay' (e.g. Martin and Whitley 2010; League of European Research Universities (LERU) 2012). There was resigned acceptance that although PREAs remain contentious, and evidence about their operation is uneven, they nevertheless are considered useful for multiple purposes. They enable governments to map, prioritize, and capitalize (better) upon research and researcher capacity within a science system. They are an accepted means to allocate research funding and infrastructure resources based upon such maps, prioritizations, and investment plans and strategies (e.g. Strehl, Reisinger and Kalatschan 2007;

European Commission 2009; Hicks 2010; Olson and Rapporteurs 2011; Organisation for Economic Co-operation and Development (OECD) 2011; Cunningham, Salavetz and Tuytens 2012; Mahieu, Arnold and Kolarz 2013; Higher Education Funding Council for England (HEFCE) 2014; Arocena, Göransson and Sutz 2018).

Literature often neither sought nor found standardization or 'best practice' of PREAs. There remain open questions, and unresolved debates, for example, how to improve design and deployment of PREA-related strategies, research funding mechanisms, performance assessment methods, key criteria, how often to conduct evaluation, whether to evaluate academic and/or non-academic research, whether to distinguish between researchers and research environments, and how to determine the most appropriate unit(s) and subject(s) of assessment (e.g. Wooding and Grant 2003; Organisation for Economic Co-operation and Development (OECD) 2009, 2010a,b; Ministry of Education 2012; Reale et al. 2018; see also Sivertsen 2017; Regan and Henchion 2019).

Despite this agnosticism regarding 'best' arrangements, there were fears of isomorphism—particularly of widespread diffusion of the UK's RAE/REF arrangements, either in entirety or specific elements, like arrangements to evaluate research 'impact'. Patterns of exploration, testing, and learning by various stakeholders (e.g. research funders, policymakers) were seen as enabling such adoption, translation, travel, and/or transplantation of PREAs from one country, region, or university context to another. Similarly, pathways were observed for 'trickle down' of national arrangements into bespoke—and sometimes highly contentious—local arrangements inside particular universities and other public research organizations (e.g. Iorwerth 2005; Mahieu, Arnold and Kolarz 2013; Ohtani, Kamo and Kobayashi 2013; Mok 2014; Aagaard 2015; Geuna and Piolatto 2016; Woelert and McKenzie 2018; see also Lind 2019).

4. Discussion

4.1 Limitations of literature on research evaluation arrangements?

Our analysis suggests five limitations across this set of PREA-related literature. *First*, there are many user-driven, policymaker/funder-commissioned reports and primarily descriptive approaches. A total of 28% of our literature set was explicitly policy/practice-oriented (i.e. policy report format) and 48% provided primarily thick descriptions of specific PREAs. Such literature is useful. However, user-oriented, thick descriptions *alone* seem insufficient to allow more critical perspectives and predictions regarding say, *effects* of arrangements and/or *reactions* (strategies, behaviours) of different organizational actors subjected to them (e.g. research funding agencies, universities, localized and more global knowledge communities). Similarly, descriptive accounts, even when oriented towards policy learning, may in fact *hinder* it because of a lack of analytical comparative foundations (and make it difficult to achieve 'mutual learning' across PREAs, as recommended by Sivertsen 2017). *Descriptions* of PREAs may make them appear comparable, transferable, or generalizable. Such comparisons are, however, often superficial. Lacking critical understanding of the use of whole or partial arrangements could lead to wide-ranging unintended and unexpected effects.

A *second* limitation is the pervasive, methodologically intractable unknowns in the literature concerning whether PREAs do produce, promote, or hinder research with specific performance-related

Table 3. Summary of limitations evident in and novel research agenda elements suggested by our critical review of PREA-related literature

Limitation	Novel research agenda element
1. Many user-driven, policymaker/funder-commissioned reports, and primarily descriptive approaches to PREAs—hindering policy learning and posing generalization issues	1. Develop and test comparative analytical frameworks to address PREAs
2. Methodologically intractable unknowns about whether PREAs improve or harm research performance-related areas, even societal impacts	2. Incorporate effectiveness concerns when researching PREAs
3. Uncertainty whether there are increasing or diminishing returns in more extensive and intrusive use of (certain kinds of) PREAs	3. Add casual attribution when studying effects of PREAs
4. Overemphasis on self-reported effects (rather than attributable change) primarily in universities and typically at the micro-level of individual researchers and their research lines in specific cases/contexts	4. Include effects on the structure of global knowledge communities and bodies of knowledge when assessing PREAs
5. Absence of comparative frameworks to account for the multifaceted aspects involved in developing and deploying PREAs in various contexts	

properties (e.g. excellence, novelty, breakthrough, long-term focus, societal relevance, or impact). This is linked to a *third* limitation; the literature is inconclusive in answering whether—particularly after seeming early gains in using certain PREAs in specific countries—there are now *increasing* or *diminishing returns* for policymakers and universities to develop and deploy seemingly ever more expensive, extensive, and potentially intrusive arrangements.

A *fourth* limitation is that research on effects of PREAs has primarily focused on (*self-*)*reported changes* in universities. Reported effects—let alone causally *attributable changes*—to *structures* and *organizations* of national, trans-national, and trans-organizational *research fields* (knowledge communities, knowledge properties) have received *much* less attention. Most research has focused upon micro-level changes to research topics or topic portfolios pursued by researchers in *specific* universities, fields, and/or national systems.

It is clear that design and deployment of PREAs does not take place in a vacuum. PREAs are parts of and are strongly ‘coupled’ to a wider universe of path-dependent, dynamic activities, and exercise of power, authority, resources, politics, and policy machinery (Whitley 2016). And yet a *fifth* limitation here is the absence of comparative frameworks to account for these aspects across the many and various development and use contexts of PREAs.

4.2 A novel research agenda on PREAs?

We believe four elements for a novel research agenda on PREAs emerge from our critical review and analysis of the state-of-the-art. First, very few, if any, *analytical* frameworks exist to study and compare research evaluation arrangements. There are examples of *comparative* frameworks (Geuna and Martin 2003) but most, with a possible exception in Whitley, Glaser and Laudel (2018) use *descriptive* not *analytical* characteristics. This reduces analytical capacity and availability of heuristics and theory to explain the many interacting mechanisms present in and across micro, meso, and macro levels of the global science system. A novel research agenda could therefore *first* include *development and testing of comparative analytical frameworks*.

Literature on rationales for PREAs has predominantly dealt with *efficiency* concerns. They determine—we argue—whether arrangements have achieved *what* they set out to achieve. Studying efficiency of research evaluation as a policy instrument is a worthy pursuit. However, there are practical and analytical limitations inherent in this delineation of a research agenda. A *second* element of a novel research agenda would be to *incorporate effectiveness*

concerns, that is, are the ‘right’ things being done in the science system? This should trace beyond *localized* conditions for research (e.g. at universities) to incorporate treatment of potential changes in the structure of *global* ‘research fields’ (c.f. Nedeva 2013).

Literature studying effects of PREAs on the science system has also largely focused on *ad hoc associations* between effects and measures. The arrangement under discussion is commonly taken to be a *universal* or *singular* enabler of the observed effects. A *third* element of a novel research agenda could be to attempt to *add causal attribution* to verify such assumptions.

Finally, we see from the literature that PREAs typically target *research organizations* in national policy and funding spaces. Correspondingly studies seemingly rarely study effects *beyond* those for universities in their own local context.¹³ A *fourth* and final element of a novel research agenda would seem to be to *include effects on the structure of global knowledge communities and bodies of knowledge*. A summary of these five limitations and four novel agenda elements is provided in Table 3.

5. Conclusion

Our critical review and purposive analysis of 354 pieces of literature we feel addresses the state-of-the-art on PREAs. It spanned works published from 1968 to 2018 and encompassed both scholarly and policy/practice-related research orientations. We believe our analysis satisfied our research aims, that is, to enable us to highlight key arguments, analyse limitations, and to suggest how to progress the research agenda in this area.

From our review we can conclude, first, analytical comparative frameworks are needed to study PREAs. Second, not only efficiency but also effectiveness concerns should be considered for PREAs. Third, studies should be devised and conducted on science system-level effects of PREAs and how global research fields are affected rather than just particular studies of local settings. Fourth, methodologies need to be advanced to measure and attribute these effects of PREAs on the (global) science system.

All four elements of this novel research agenda seem both necessary and challenging. There are numerous levels of mediation of effects and inherent complexities to unpack layer upon layer of research-related conditions here. We limited our article’s aims to (re-)opening the research agenda on PREAs and their effects on the science system by means of a critical, purposive, inductive examination of PREA-related research themes, and identification of agenda

gaps. Developing analytical frameworks for PREAs, perhaps even outlining ‘ideal’ types of PREAs, and stretching studies of effects to include research fields appear essential. Similarly, learning to cope better with effectiveness, measurement and attribution issues seem necessary next steps, to take studies of PREAs further, to the benefit of both academic and practitioner interests.

Notes

1. We provide our full definition of ‘performance-based research evaluation arrangements’ in the following section, and distinguish it from research evaluation ‘systems’.
2. We focused our attention on journals publishing on topics of higher education studies, higher education policy, higher education management, sociology of science and science and technology policy studies, as well as fields like health policy and studies where research evaluation is addressed as a side issue in larger discussions (e.g. on priority setting). Indicative journals in our literature set include *Cambridge Journal of Economics, Education Policy, Evidence & Policy, Higher Education, Higher Education Quarterly, Journal of Electronic Resources in Medical Libraries, Journal of Higher Education Policy and Management, Journal of Infometrics, Journal of Information Science, Journal of Sociology of Education, Journal of the Association for Information Science and Technology, Management and Policy, Minerva, Policy and Society, Political Studies Review, Public Administration, Public Management Review, Research Evaluation, Research Policy, Science, Science and Public Policy, Scientometrics, and Tertiary Education and Management*.
3. All literature was in English except for one piece in Spanish. We did not try to access private, commercially sensitive, or confidential evaluations of specific research performers or funders. The entire set of academic and grey literature is heterogeneous, even though we confined our search to publicly available, English-language materials. This is likely due to significant involvement of funders in sponsoring research say, to audit their resource allocation processes and evaluate the outcomes of their funded research.
4. Our full country coverage includes Australia, Austria, Belgium/Flanders, Brazil, Bulgaria, Canada, China, Czech Republic, Denmark, Estonia, Finland, Germany, Hong Kong, Hungary, India, Ireland, Italy, Japan, Korea, Latvia, Lithuania, Mexico, Morocco, Netherlands, New Zealand, Norway, Poland, Romania, Slovak Republic, Slovenia, South Africa, Spain, Sweden, Switzerland, UK, USA, and Uruguay.
5. Correspondingly we cite some literature in multiple theme sections of our later findings, when their *secondary* message(s) are relevant, denoted by ‘see also’ in our citations. Our choice to allocate by primary theme rather than cluster in multiple themes by coverage of all issues is of course contentious. However, we believe this subjective approach provides a more useful thematic clustering for our purposes than exhaustively cataloguing by primary, secondary, tertiary, etc. themes.
6. We considered coding our subjective judgement of the apparent quality of literature. We decided against this step, in case it influenced our later analysis.
7. For ease of reference, we also included in our database columns for author name(s), title of the work, and publication year.

8. We provide numbers indicatively to show how much literature clustered into each theme, and the surveyed balance of approaches and content (e.g. scholarly, policy/practice-orientations). Our numbers and percentages do not constitute general impressions about the broader universe of evaluation-related research that exists outside our specific analytical boundaries for literature on PREAs.
9. One piece of literature we coded ‘other’; it was more abstract in its descriptive approach.
10. We considered ‘rationales’ for research evaluation to be within our analytical remit because they were present in the PREA-related literature. Our inductive clustering of themes reflects that these issues were being discussed in material within the scope of our PREA definition.
11. Of all the clustering themes, literature in Theme 3 was the most what we would call ‘synthetic’, in that primary messages often combined aspects of one or more of our analytical themes.
12. This ‘strategic approach’ concerns *selectivity* and *concentration* of research resources to research areas, researchers and teams, and universities displaying characteristics associated with excellence: share of highly cited publications, citations/impact, external grants capture, industry links, and patents. New tools and data to measure this notion of excellence have been associated with pressures on research systems to adapt to dominant ideas around value for money, steering and control, accountability, and measurement (Butler 2003; see also Debackere and Glanzel 2003; Geuna and Martin 2003; Linkova 2014).
13. Other effects considered do include researcher careers, but predominantly just ‘organizational’ careers, still constraining analysis within the policy/funding/university ‘space’ and not on to ‘cognitive’ or ‘knowledge community’ careers in research fields (c.f. terminology from Laudel 2017).

Acknowledgements

We are grateful to two anonymous referees for their insightful feedback.

Funding

This work was supported by the Swedish Foundation for Social Science and Humanities Research (Riksbankens Jubileumsfond) [FSK15-0881:1]. Knowledge in science and policy: creating an evidence base for converging modes of governance in policy and science (KNOWSCIENCE).

Conflict of interest statement. None declared.

References

- Aagaard, K. (2015) ‘How Incentives Trickle Down: Local Use of a National Bibliometric Indicator System’, *Science and Public Policy*, 42: 725–37.
- Aagaard, K., Bloch, C., and Schneider, J. W. (2015) ‘Impacts of Performance-Based Research Funding Systems: The Case of the Norwegian Publication Indicator’, *Research Evaluation*, 24: 106–17.
- Abramo, G., Cicero, T., and D’Angelo, A. C. (2013) ‘National Peer-Review Research Assessment Exercises for the Hard Sciences Can Be a Complete Waste of Money: The Italian Case’, *Scientometrics*, 95: 311–24.
- Abramo, G., and D’Angelo, C. A. (2011) ‘Evaluating Research: From Informed Peer Review to Bibliometrics’, *Scientometrics*, 87: 499–514.

- Abramo, G., D'Angelo, C. A., and Di Costa, F. (2008) 'Assessment of Sectoral Aggregation Distortion in Research Productivity Measurements', *Research Evaluation*, 17: 111–21.
- Abramo, G., D'Angelo, C. A., and Di Costa, F. (2011) 'National Research Assessment Exercises: The Effects of Changing the Rules of the Game during the Game', *Scientometrics*, 88: 229–38.
- Adams, J., and Gurney, K. (2010) *Funding Selectivity, Concentration and Excellence—How Good Is the UK's Research?* <http://www.rin.ac.uk/system/files/attachments/Funding_selectivity_concentration_excellence_-_Exec_Summ.pdf> accessed 28 Mar 2019.
- Adler, P. A., and Adler, P. (1987) *Membership Roles in Field Research* (SAGE University Paper Series on Qualitative Research, Vol. 6). Newbury Park, CA: SAGE.
- Agyemang, G., and Broadbent, J. (2015) 'Management Control Systems and Research Management in Universities. An Empirical and Conceptual Exploration', *Accounting, Auditing & Accountability Journal*, 28: 1018–46.
- Ancaiani, A., Anfossi, A. F., Barbara, A., Benedetto, S., Blasi, B., Carletti, V., Cicero, T., Ciolfi, A., Costa, F., Colizza, G., Costantini, M., di Cristina, F., Ferrara, A., Lacatena, R. M., Malgarini, M., Mazzotta, I., Nappi, C. A., Romagnosi, S., and Sileoni, S. (2015) 'Evaluating Scientific Research in Italy: The 2004–10 Research Evaluation Exercise', *Research Evaluation*, 24: 242–55.
- Arnold, E., and Mahieu, B. (2015) *R&D Evaluation Methodology and Funding Principles. Summary Report*. Technopolis. <<https://www.slide-share.net/ipnmetodika/2104-summary-report>> accessed 28 Mar 2019.
- Arnold, E., et alFarla, K., Kolarz, P., Mahieu, B., and Peter, V. (2014) *The Role of Metrics in Performance-Based Research Funding Systems. A Report to the Russell Group*. Brighton, UK: Technopolis Group.
- Arocena, R., Göransson, B., and Sutz, J. (2018) 'Towards Making Research Evaluation More Compatible with Developmental Goals', *Science and Public Policy*, 46: 210–8.
- Auranen, O., and Nieminen, M. (2010) 'University Research Funding and Publication Performance—An International Comparison', *Research Policy*, 39: 822–34.
- Barker, K. (2007) 'The UK Research Assessment Exercise: The Evolution of a National Research Evaluation System', *Research Evaluation*, 16: 3–12.
- Bence, V., and Oppenheim, C. (2005) 'The Evolution of the UK's Research Assessment Exercise: Publications, Performance and Perceptions', *Journal of Educational Administration and History*, 37: 137–55.
- Benner, M., and Sandström, U. (2000) 'Institutionalizing the Triple Helix: Research Funding and Norms in the Academic System', *Research Policy*, 29: 291–301.
- Bertocchi, G., Gambardella, A., Jappelli, T., Nappi, C. A., and Peracchi, F. (2015) 'Bibliometric Evaluation vs. Informed Peer Review: Evidence from Italy', *Research Policy*, 44: 451–66.
- Bianco, M., Gras, N., and Sutz, J. (2016) 'Academic Evaluation: Universal Instrument? Tool for Development?', *Minerva*, 54: 399–421.
- Bozeman, B., Dietz, J., and Gaughan, M. (2001) 'Scientific and Technical Human Capital: An Alternative Model for Research Evaluation', *International Journal of Technology Management*, 7: 716–40.
- Bozeman, B., and Sarewitz, D. (2011) 'Public Value Mapping and Science Policy Evaluation', *Minerva*, 49: 1–23.
- Broadbent, J. (2010) 'The UK Research Assessment Exercise: Performance Measurement and Resource Allocation', *Australian Accounting Review*, 52: 14–23.
- Butler, L. (2003) 'Explaining Australia's Increased Share of ISI Publications—The Effects of a Funding Formula Based on Publication Counts', *Research Policy*, 32: 143–55.
- Butler, L. (2007) 'Assessing University Research: A Plea for a Balanced Approach', *Science and Public Policy*, 34: 565–74.
- Butler, L. (2008) 'Using a Balanced Approach to Bibliometrics: Quantitative Performance Measures in the Australian Research Quality Framework', *Ethics in Sciences and Environmental Politics*, 8: 83–92.
- Campbell, M., and Boxall, M. (2004) *Better Accountability Revisited: Review of Accountability Costs 2004*. PA Consulting Group. <https://dera.ioe.ac.uk/4985/1/rd06_04.pdf> accessed 28 Mar 2019.
- Canibano, C., Vilardell, I., Corona, C., and Benito-Amat, C. (2018) 'The Evaluation of Research Excellence and the Dynamics of Knowledge Production in the Humanities: The Case of History in Spain', *Science and Public Policy*, 45: 775–89.
- Cattaneo, M., Meoli, M., and Signori, A. (2016) 'Performance-Based Funding and University Research Productivity: The Moderating Effect of University Legitimacy', *The Journal of Technology Transfer*, 41: 85–104.
- Claeys-Kulik, A. L., and Estermann, T. (2015) *Define Thematic Report: Performance-Based Funding of Universities in Europe*. European University Association. <<https://eua.eu/downloads/publications/define%20thematic%20report%20performance-based%20funding%20of%20universities%20in%20europe.pdf>> accessed 28 Mar 2019.
- Court, S. (1999) 'Negotiating the Research Imperative: The Views of UK Academics on Their Career Opportunities', *Higher Education Quarterly*, 53: 65–87.
- Cozzens, S. E. (1981) 'Taking the Measure of Science: A Review of Citation Theories', *Newsletter of the International Society for the Sociology of Knowledge*, 7: 16–21.
- Cruz-Castro, L., and Sanz-Menéndez, L. (2007). 'Research Evaluation in Transition: Individual versus Organisational Assessment in Spain', in Whitley Richard and Gläser Jochen (eds) *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems*, pp. 205–23. Dordrecht: Kluwer.
- Cunningham, P., Salavetz, A., and Tuytens, P. (2012) *Monitoring Social Sciences and Humanities*. Metris. Synthesis Report 2012. Brighton, UK: Technopolis Group.
- De Boer, H., Jongbloed, B. W. A., Benneworth, P. S., Cremonini, L., Kolster, R., Kottmann, A., Lemmens-Krug, K., and Vossensteyn, J. J. (2015) *Performance-Based Funding and Performance Agreements in Fourteen Higher Education Systems. Report for the Ministry of Education, Culture and Science*. Netherlands: Centre for Higher Education Policy Studies, Universiteit Twente.
- De Jong, S., van Arensbergen, P., Daemen, F. F., van der Meulen, B., and van den Besselaar, P. (2011) 'Evaluation of Research in Context: An Approach and Two Cases', *Research Evaluation*, 20: 61–72.
- De Rijcke, S., Wouters, P. F., Rushforth, A. D., Franssen, T. P., and Hammarfelt, B. (2016) 'Evaluation Practices and Effects of Indicator Use—A Literature Review', *Research Evaluation*, 25: 161–9.
- Debackere, K., and Glanzel, W. (2003) *Using a Bibliometric Approach to Support Research Policy Decisions: The Case of the Flemish BOF-Key*. Research Report 0306, D/2003/2376/06: 1–26. Leuven, Belgium: Katholieke Universiteit Leuven.
- Demiris, G., Oliver, D. P., and Washington, K. T. (2019) 'Defining and Analyzing the Problem', in Demiris, G., Oliver, D. P., and Washington, K. T. (eds) *Behavioral Intervention Research in Hospice and Palliative Care*, pp. 27–39. Amsterdam, Netherlands: Elsevier Science and Technology Academic Press.
- Donovan, C. (2007) 'The Qualitative Future of Research Evaluation', *Science and Public Policy*, 34: 585–97.
- Donovan, C. (2008) 'The Australian Research Quality Framework: A Live Experiment in Capturing the Social, Economic, Environmental, and Cultural Returns of Publicly Funded Research', in Coryn, C. and Scriven, M. (eds) *Reforming the Evaluation of Research. New Directions for Evaluation*, pp. 47–60. San Francisco, CA: Jossey-Bass.
- Donovan, C. (2009) 'Gradgrinding the Social Sciences: The Politics of Metrics of Political Science', *Political Studies Review*, 7: 73–83.
- Elsevier (2011) *International Comparative Performance of the UK Research Base. A Report Prepared for the Department of Business, Innovation and Skills*. <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/32489/11-p123-international-comparative-performance-uk-research-base-2011.pdf> accessed 28 Mar 2019.
- Elsevier (2013) *International Comparative Performance of the UK Research Base. A Report Prepared for the Department of Business, Innovation and*

- Skills. <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/263729/bis-13-1297-international-comparative-performance-of-the-UK-research-base-2013.pdf> accessed 28 Mar 2019.
- Elton, L. (2000) 'The UK Research Assessment Exercise: Unintended Consequences', *Higher Education Quarterly*, 54: 274–83.
- European Centre for Strategic Management of Universities (ESMU) (2010) *Funding Higher Education: A View across Europe*. <https://uniko.ac.at/modules/download.php?key=4488_DE_O&cs=8E8> accessed 28 Mar 2019.
- European Commission (2009) *The Role of Community Research Policy in the Knowledge-Based Economy. Experts Group Report*. Belgium. <<https://publications.europa.eu/en/publication-detail/-/publication/8edc4431-c8e7-49ba-a98a-ba0dc911ca13/language-en>> accessed 28 Mar 2019.
- Eyre-Walker, A., and Stoletzki, N. (2013) 'The Assessment of Science: The Relative Merits of Post-Publication Review, the Impact Factor, and the Number of Citation', *PLoS Biology*, 11: e1001675.
- Ferlie, E., Musselin, C., and Andresani, G. (2008) 'The Steering of Higher Education Systems: A Public Management Perspective', *Higher Education*, 56: 325–48.
- Fiala, D. (2013) 'Science Evaluation in the Czech Republic: The Case of Universities', *Societies*, 3: 1–13.
- Franceschet, M., and Costantini, A. (2011) 'The First Italian Research Assessment Exercise: A Bibliometric Perspective', *Journal of Infometrics*, 5: 275–91.
- Frankel, M., Goddard, A., and Ransow, G. (2014) 'Golden Triangle Pulls Ahead in REF Shake-out. UCL and KCL Ascend Power Rankings, Manchester and Leeds Fall', *Research Fortnight, Issue*, 4470: 1–20.
- Franzoni, C., Scellato, G., and Stephan, P. (2011) 'Science Policy: Changing Incentives to Publish', *Science*, 333: 702–3.
- Frølich, N. (2008) *The Politics of Steering by Numbers: Debating Performance-Based Funding in Europe*. Oslo, Norway: NIFU STEP.
- Frølich, N. (2011) 'Multi-Layered Accountability. Performance-Based Funding of Universities', *Public Administration*, 89: 840–59.
- Frølich, N., Schmidt, E., and Rosa, M. J. (2010) 'Funding Systems for Higher Education and Their Impacts on Institutional Strategies and Academia. A Comparative Perspective', *International Journal of Educational Management*, 24: 7–21.
- Galleron, I., Ochsner, M., Spaapen, J., and Williams, G. (2017) 'Valorizing SSH Research: Towards a New Approach to Evaluate SSH Research' Value for Society', *Fteval Journal for Research and Technology Policy Evaluation*, 44: 35–41.
- Geuna, A., and Martin, B. R. (2003) 'University Research Evaluation and Funding: An International Comparison', *Minerva*, 41: 277–304.
- Geuna, A., and Piolatto, M. (2016) 'The Development of Research Assessment in the UK and Italy: Costly and Difficult, but Probably Worth (for a While)', *Research Policy*, 45: 260–71.
- Gibson, J., Anderson, D. L., and Tressler, J. (2014) 'Which Journal Rankings Best Explain Academic Salaries? Evidence from the University of California', *Economic Inquiry*, 52: 1322–40.
- Gläser, J. (2007) 'The Social Orders of Research Evaluation Systems', in Whitley Richard and Gläser Jochen (eds) *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems*, pp. 245–66. Dordrecht: Springer.
- Gläser, J. (2019) 'How Can Governance Change Research Content? Linking Social Policy Studies to the Sociology of Science', in Simon, D., Kuhlmann S., Stamm, J., and Canzler, W. (eds) *Handbook on Science and Public Policy*, pp. 419–47. Cheltenham, UK and Northampton, MA, USA: Edward Elgard.
- Gläser, J., Lange, S., Laudel, G., and Schimank, U. (2010) 'Informed Authority? The Limited Use of Research Evaluation Systems for Managerial Control in Universities', in Whitley, R., Gläser, J., and Engwall, L. (eds) *Reconfiguring Knowledge Production: Changing Authority Relationships in the Sciences and Their Consequences for Intellectual Innovation*, pp. 149–369. Oxford: Oxford University Press.
- Grant, J. (2010) *Capturing Research Impacts. A Review of International Practice*. RAND Europe. <https://www.rand.org/content/dam/rand/pubs/documented_briefings/2010/RAND_DB578.pdf> accessed 28 Mar 2019.
- Greenhalgh, T., and Fahy, N. (2015) 'Research Impact in the Community-Based Health Sciences: An Analysis of 162 Case Studies from the 2014 UK Research Excellence Framework', *BMC Medicine*, 13: 1–12.
- Hamann, J. (2016) 'The Visible Hand of Research Performance Assessment', *Higher Education*, 72: 761–79.
- Hammarfelt, B., and de Rijcke, S. (2015) 'Accountability in Context: Effects of Research Evaluation Systems on Publication Practices, Disciplinary Norms, and Individual Working Routines in the Faculty of Arts at Uppsala University', *Research Evaluation*, 24: 63–77.
- Hare, P. G. (2003) 'The United Kingdom's Research Assessment Exercise: Impact on Institutions, Departments, Individuals', *Higher Education Management and Policy*, 15: 43–62.
- Harland, T., Tidswell, T., Everett, D., Hale, L. and Pickering, N. (2010) 'Neoliberalism and the Academic as Critic and Conscience of Society', *Teaching in Higher Education*, 15: 185–96.
- Higher Education Funding Council for England (HEFCE) (2014) *A Review of QR Funding in English HEIs. Process and Impact*. <https://www.praxisauril.org.uk/sites/praxisuniko.org.uk/files/HEFCE_2014_qrreview.pdf> accessed 28 Mar 2019.
- Henkel, M. (1999) 'The Modernisation of Research Evaluation: The Case of the UK', *Higher Education*, 38: 105–22.
- Hicks, D. (2010) 'Overview of Models of Performance-Based Research Funding Systems', in OECD, *Performance Based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings*, 1 December, pp. 23–52. Paris, France: OECD Publishing.
- Hicks, D. (2012) 'Performance-Based University Research Funding Systems', *Research Policy*, 41: 251–61.
- Himanen, L., Auranen, O., Puuska, H.-M., and Nieminen, M. (2009) 'Influence of Research Funding and Science Policy on University Research Performance: A Comparison of Five Countries', *Science and Public Policy*, 36: 419–30.
- Holmes, R. (2015) 'My Tongue on Your Theory: The Bittersweet Reminder of Every-Thing Unnameable', *Discourse, Studies in the Cultural Politics of Education*, 37: 662–79.
- Hug, S. E., Ochsner, M., and Daniel, H.-D. (2013) 'Criteria for Assessing Research Quality in the Humanities: A Delphi Study among Scholars of English Literature, German Literature and Art History', *Research Evaluation*, 22: 369–83.
- Hughes, A., Kitson, M., and Bullock, A. (2013) *The Dual Funding Structure for Research in the UK: Research Council and Funding Council Allocation Methods and the Pathways to Impact of UK Academics*. Research report issued 5 April. Cambridge and London, UK: University of Cambridge and Imperial College London.
- Ioannidis, J. (2018) 'Meta-Research: Why Research on Research Matters', *PLoS Biology*, 16: e2005468.
- Iorwerth, A. (2005) 'Methods of Evaluating University Research around the World'. Working Paper 2005-04. Department of Finance, Canada.
- Johnston, R. (1994) 'Effects of Resource Concentration on Research Performance', *Higher Education*, 28: 25–37.
- Jonkers, K., and Sachwald, F. (2018) 'The Dual Impact of 'Excellent' Research on Science and Innovation: The Case of Europe', *Science and Public Policy*, 45: 159–74.
- Jonkers, K., and Zacharewicz, T. (2015) *Performance Based Funding: A Comparative Assessment of Their Use and Nature in EU Member States*. EUR 27477. Brussels, Belgium: European Union JRC Science Hub, JRC97684.
- Kallerud, E., Finnbjörnsson, T., Geschwind, L., Häyrinen-Alestalo, M., Ramberg, I., Siune, K., and Tuominen, T. (2011) *Public Debate on Research Policy in the Nordic Countries. A Comparative Analysis of Actors and Issues (1998–2007)*. <<http://www.nifu.no/Norway/Publications/2011/Webrapport%2011-2011.pdf>> accessed 28 Mar 2019.
- Kaufmann, A., and Kasztler, A. (2009) 'Differences in Publication and Dissemination Practices between Disciplinary and Transdisciplinary Science

- and the Consequences for Research Evaluation', *Science and Public Policy*, 36: 215–27.
- Kenna, R., and Berche, B. (2011) 'Normalization of Peer-Evaluation Measures of Group Research Quality across Academic Disciplines', *Research Evaluation*, 20: 107–16.
- Knowles, C., and Burrows, R. (2014) 'The Impact of Impact', *Etnográfica*, 18: 237–54.
- Kwok, J. T. (2013) *Impact of ERA Research Assessment on University Behaviour and Their Staff*. National Tertiary Education Union. <<https://core.ac.uk/download/pdf/30677048.pdf>> accessed 28 Mar 2019.
- Langfeldt, L., Nedeva, M., Sörlin, S., and Thomas, D. A. (2020) 'Co-Existing Notions of Research Quality: A Framework to Study Context-Specific Understanding of Good Research', *Minerva*, 58: 115–137.
- Laudel, G. (2005) 'Quality-Only Assumption. Is External Research Funding a Valid Indicator for Research Performance?', *Research Evaluation*, 14: 27–34.
- Laudel, G. (2017) 'How Do National Career Systems Promote or Hinder the Emergence of New Research Lines?', *Minerva*, 55: 341–69.
- Lee, F. S., and Harley, S. (1998) 'Peer Review, the Research Assessment Exercise and the Demise of Non-Mainstream Economics', *Capital & Class*, 66: 23–51.
- Lee, F. S., Pham, X., and Gu, G. (2013) 'The UK Research Assessment Exercise and the Narrowing of UK Economics', *Cambridge Journal of Economics*, 37: 693–717.
- Leisyte, L., and Westerheijden, D. F. (2014) *Research Evaluation and Its Implications for Academic Research in the United Kingdom and the Netherlands*. Discussion paper No. 1-2014. Dortmund, Germany: Technische Universität Dortmund.
- Lepori, B., Masso, J., Jablecka, J., Sima, K., and Ukrainski, K. (2009) 'Comparing the Organization of Public Research Funding in Central and Eastern European Countries', *Science and Public Policy*, 36: 667–81.
- League of European Research Universities (LERU) (2012) *Research Universities and Research Assessment*. <<https://www.leru.org/files/Research-Universities-and-Research-Assessment-Full-paper.pdf>> accessed 28 Mar 2019.
- Lind, J. (2019) 'The Missing Link: How University Managers Mediate the Impact of a Performance-Based Research Funding System', *Research Evaluation*, 28: 84–93.
- Linkova, M. (2014) 'Unable to Resist: Researchers' Responses to Research Assessment in the Czech Republic', *Human Affairs*, 24: 78–88.
- Luo, J., Ordóñez-Matamoros, G., and Kuhlmann, S. (2019) 'The Balancing Role of Evaluation Mechanisms in Organizational Governance—The Case of Publicly Funded Research Institutions', *Research Evaluation*, 28: 344–54.
- Luukkonen, T. (1997) 'Why Has Latour's Theory of Citations Been Ignored by the Bibliometric Community? Discussion of Sociological Interpretations of Citation Analysis', *Scientometrics*, 38: 27–37.
- Luukkonen, T., and Thomas, D. A. (2016) 'The 'Negotiated Space' of University Researchers' Pursuit of a Research Agenda', *Minerva*, 54: 99–127.
- Macdonald, S., and Kam, J. (2010) 'Counting Footnotes: Citability in Management Studies', *Scandinavian Journal of Management*, 26: 189–203.
- Mahieu, B., Arnold, E., and Kolarz, P. (2013) *Measuring Scientific Performance for Improved Policy Making. Literature Review*. Brighton, UK: Technopolis Group.
- Mahieu, B., Arnold, E., and Kolarz, P. (2014) *Measuring Scientific Performance for Improved Policy Making. Final Report—Summary (D6)*. Brighton, UK: Technopolis Group.
- Mahieu, B., and Arnold, E. (2015) *R&D Evaluation Methodology and Funding Principles. Final Report 1: The R&D Evaluation Methodology*. Brighton, UK: Technopolis Group.
- Mahieu, B., Brown, N., Fridholm, T., Giarracca, F., Hinojosa, C., Horvath, A., Potau, X., Quas, A., and Tummers, A et al. (2013) *Measuring Scientific Performance for Improved Policy-Making. Current Practice in the EU Member States (D3—Survey Report)*. Brighton, UK: Technopolis Group.
- Martin, B., and Whitley, R. (2010) 'The UK Research Assessment Exercise: A Case of Regulatory Capture?', in Whitley, R., Glaser, J., and Engwall, L. (eds) *Reconfiguring Knowledge Production: Changing Authority Relationships in the Sciences and Their Consequences for Intellectual Innovation*, pp. 51–80. Oxford, UK: Oxford University Press.
- Martin, B. R. (2011) 'The Research Excellence Framework and the 'Impact Agenda': Are we Creating a Frankenstein Monster?', *Research Evaluation*, 20: 247–54.
- Martin, B. R. (2016) 'What is Happening to Our Universities?' SPRU Working Paper Series 2016-03, 1-26. Brighton, UK: SPRU. ISSN 2057-6668.
- Merton, R. K. (1968) 'The Matthew Effect in Science. The Reward and Communication Systems of Science Are Considered', *Science*, 159: 56–63.
- Mingers, J., and White, L. (2015) *Throwing Out the Baby with the Bathwater: The Undesirable Effects of National Research Assessment Exercises on Research*. <<https://arxiv.org/ftp/arxiv/papers/1502/1502.00658.pdf>> accessed 28 Mar 2019.
- Ministry of Education (2012) *An International Comparison of Performance-Based Research Funding Systems (PBRFS)*. Wellington, New Zealand: New Zealand Parliament Library.
- Mok, K. H. (2014) 'Enhancing Quality of Higher Education for World-Class Status', *Chinese Education & Society*, 47: 44–64.
- Morris, N. (2010) 'Authority Relations as Conditions for, and Outcome of, Shifts of Governance: The Limited Impact of the UK Research Assessment Exercise on the Biosciences', in Whitley, R., Glaser, J., and Engwall, L. (eds) *Reconfiguring Knowledge Production: Changing Authority Relationships in the Sciences and Their Consequences for Intellectual Innovation*, pp. 239–64. Oxford, UK: Oxford University Press.
- Murphy, T., and Sage, D. (2014) 'Perceptions of the UK's Research Excellence Framework 2014: A Media Analysis', *Journal of Higher Education Policy and Management*, 36: 603–15.
- Nedeva, M. (2013) 'Between the Global and the National: Organising European Science', *Research Policy*, 42: 220–30.
- Nedeva, M., Braun, D., Edler, J., Glaser, J., Laredo, P., Laudel, G., Luukkonen, T., Stampfer, M., Thomas, D., and Whitley, R. (2012) *Understanding and Assessing the Impact and Outcomes of the ERC and Its Funding Schemes, EURECIA Final Synthesis Report*. <http://erc.europa.eu/sites/default/files/document/file/eurecia_final_synthesis_report.pdf> accessed 28 Mar 2019.
- Neufeld, J., and von Ins, M. (2011) 'Informed Peer Review and Uninformed Bibliometrics?', *Research Evaluation*, 20: 31–46.
- Ochsner, M., Hug, S. E., and Daniel, H.-D. (2012) 'Indicators for Research Quality for Evaluation of Humanities Research: Opportunities and Limitations', *Bibliometrie - Praxis und Forschung*, 4: 1–17.
- Ohtani, R., Kamo, M., and Kobayashi, N. (2013) 'An Introduction to the Research Excellence Framework: A New Research Evaluation Framework for Universities in the UK—A Comparison with the Status of Research Evaluation in Japan', *Synthesiology*, 6: 118–25.
- Olson, S., and Rapporteurs, S. M. (2011) *Measuring the Impacts of Federal Investments in Research. A Workshop Summary*. USA: National Academy of Sciences.
- Olssen, M. (2016) 'Neoliberal Competition in Higher Education Today: Research, Accountability and Impact', *British Journal of Sociology of Education*, 37: 129–48.
- Organisation for Economic Co-operation and Development (OECD) (2009) *Enhancing Public Research Performance through Evaluation, Impact Assessment and Priority Setting*. <<https://www.oecd.org/sti/inn/enhancing-public-research-performance.pdf>> accessed 28 Mar 2019.
- Organisation for Economic Co-operation and Development (OECD) (2010a) *Performance-Based Funding for Public Research in Tertiary Education Institutions*. Australia: Web annex: additional country detail.
- Organisation for Economic Co-operation and Development (OECD) (2010b) *Performance-Based Funding for Public Research in Tertiary Education Institutions Workshop Proceedings*. <https://www.oecd-ilibrary.org/fr/education/performance-based-funding-for-public-research-in-tertiary-education-institutions_9789264094611-en> accessed 28 Mar 2019.

- Organisation for Economic Co-operation and Development (OECD) (2011) *Issue Brief: Public Sector Research Funding*. <<http://www.oecd.org/innovation/policyplatform/48136600.pdf>> accessed 28 Mar 2019.
- Orr, D. (2004) 'Research Assessment as an Instrument for Steering Higher Education. A Comparative Study', *Journal of Higher Education Policy and Management*, 26: 345–62.
- Otley, D. (2010) 'Research Assessment in the UK: An Overview of 1992–2008', *Australian Accounting Review*, 52: 3–13.
- PA Consulting Group (2008) *RAE 2008 Accountability Review*. London, UK: PA Consulting.
- Paradise, C., and Thoenig, J.-C. (2015) *Search of Academic Quality*. Houndmills, Basingstoke: Palgrave Macmillan.
- Reale, E., Avramov, D., Canhial, K., Donovan, C., Flecha, R., Holm, P., Larkin, C., Lepori, B., Mosoni-Fried, J., Oliver, E., Primeri, E., Puigvert, L., Scharnhorst, A., Schubert, A., Soler, M., Soòs, S., Sordé, T., Travis, C., and Van Horik, R. (2018) 'A Review of Literature on Evaluating the Scientific, Social and Political Impact of Social Sciences and Humanities Research', *Research Evaluation*, 27: 298–308.
- Rebora, G., and Turri, M. (2013) 'The UK and Italian Research Assessment Exercises Face to Face', *Research Policy*, 42: 1657–66.
- Regan, A., and Henchion, M. (2019) 'Making Sense of Altmetrics: The Perceived Threats and Opportunities for Academic Identity', *Science and Public Policy*, 46: 479–89.
- Rip, A., and van der Meulen, B. (1996) 'The Post-Modern Research System', *Science and Public Policy*, 23: 343–52.
- Roberts, P. (2007) 'Neoliberalism, Performativity and Research', *Review of Education*, 53: 349–65.
- Sandström, U., and Van den Besselaar, P. (2018) 'Funding, Evaluation, and the Performance of National Research Systems', *Journal of Infometrics*, 12: 365–84.
- Sastry, T., and Bekhradnia, B. (2014) *Using Metrics to Allocate Research Funds. A Short Evaluation of Alternatives to the Research Assessment Exercise*. Higher Education Policy Institute. <<https://www.hepi.ac.uk/wp-content/uploads/2014/02/23RAEandmetricsfullreport.pdf>> accessed 28 Mar 2019.
- Sivertsen, G. (2017) 'Unique, but Still Best Practice? The Research Excellence Framework (REF) from an International Perspective', *Palgrave Communications*, 3: 17078.
- Smith, S., Ward, V., and House, A. (2011) 'Impact' in the Proposals for the UK's Research Excellence Framework: Shifting the Boundaries of Academic Autonomy', *Research Policy*, 40: 1369–79.
- Sørensen, M. P., Bloch, C., and Young, M. (2015) 'Excellence in the Knowledge-Based Economy: From Scientific to Research Excellence', *European Journal of Higher Education*, 6: 217–36.
- Sörlin, S. (2007) 'Funding Diversity: Performance-Based Funding Regimes as Drivers of Differentiation in Higher Education Systems', *Higher Education Policy*, 20: 413–40.
- Sousa, S. B., and Brennan, J. L. (2014) 'The UK Research Excellence Framework and the Transformation of Research Production', in Musselin C. and Teixeira P. (eds) *Reforming Higher Education*, pp. 65–80. Dordrecht: Springer.
- Spaapen, J., and Van Drooge, L. (2011) 'Introducing "Productive Interactions" in Social Impact Assessment', *Research Evaluation*, 20: 211–8.
- Strehl, F., Reisinger, S., and Kalatschan, M. (2007) 'Funding Systems and their Effects on Higher Education Systems', OECD Education Working Paper, No. 6. Paris, France: OECD Publishing.
- Tahar, S., and Boutellier, R. (2013) 'Resource Allocation in Higher Education in the Context of New Public Management', *Public Management Review*, 15: 687–711.
- Technopolis (2009) *Identification and Dissemination of Lessons Learned by Institutions Participating in the Research Excellence Framework (REF) Bibliometrics Pilot. Results of the Round One Consultation*. Brighton, UK: Technopolis.
- Technopolis (2010) *REF Research Impact Pilot Exercise Lessons-Learned Project: Feedback on Pilot Submissions. Final Report*. Brighton, UK: Technopolis.
- Van der Meulen, B. (1998) 'Science Policies as Principal-Agent Games Institutionalization and Path Dependency in the Relation between Government and Science', *Research Policy*, 27: 397–414.
- Van Drooge, L., de Jong, S., Faber, M., and Westerheijden, D. (2013) *Twenty Years of Research Evaluation*. The Rathenau Instituut. <https://www.rathenau.nl/sites/default/files/2018-05/Twenty_years_of_research_evaluation_-_Rathenau_01.pdf> accessed 28 Mar 2019.
- Vincent, A. (2015) 'The Ideological Context of Impact', *Political Studies Review*, 13: 474–84.
- Waitere, H. J., Wright, J., Tremaine, M., Brown, S., and Pause, C. J. (2011) 'Choosing Whether to Resist or Reinforce the New Managerialism: The Impact of Performance-Based Research Funding on Academic Identity', *Higher Education Research & Development*, 30: 205–17.
- Wang, J., and Hicks, D. (2012) 'Policy screening by structural change detection: can policies effectively boost research system performance?', Proceedings of 17th International Conference on Science and Technology Indicators (2012). <http://works.bepress.com/diana_hicks/35/> accessed 28 Mar 2019.
- Watermeyer, R. (2014) 'Issues in the Articulation of "Impact": The Responses of UK Academics to "Impact" as a New Measure of Research Assessment', *Studies in Higher Education*, 39: 359–77.
- Watermeyer, R. (2016) 'Impact in the REF: Issues and Obstacles', *Studies in Higher Education*, 41: 199–214.
- Whitley, R. (2011) 'Changing Governance and Authority Relations in the Public Sciences', *Minerva*, 49: 359–85.
- Whitley, R. (2016) 'Varieties of Scientific Knowledge and Their Contributions to Dealing with Policy Problems: A Response to Richard Nelson's "the Sciences Are Different and the Differences Matter"', *Research Policy*, 45: 1702–7.
- Whitley, R., and Gläser, J. (eds) (2007) *The Changing Governance of the Sciences. The Advent of Research Evaluation Systems*. The Netherlands: Springer.
- Whitley, R., Glaser, J., and Laudel, G. (2018) 'The Impact of Changing Funding and Authority Relationships on Scientific Innovations', *Minerva*, 56: 109–34.
- Wilczynski, S. M. (2017) 'Other Sources of Evidence', in Wilczynski, S. M. (ed) *Critical Specialities—Treating Autism & Behavioral Challenge*, pp. 13–19. Amsterdam, Netherlands: Elsevier Science and Technology Academic Press.
- Wilsdon, J. R., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R. A. L., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J., and Johnson, B. (2015) *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. DOI: 10.13140/RG.2.1.4929.1363.
- Woelert, P. (2015) 'The "Logic of Escalation" in Performance Measurement: An Analysis of the Dynamics of a Research Evaluation System', *Policy and Society*, 34: 75–85.
- Woelert, P., and McKenzie, L. (2018) 'Follow the Money? How Australian Universities Replicate National Performance-Based Funding Mechanisms', *Research Evaluation*, 27: 184–95.
- Wooding, S., and Grant, J. (2003) *Assessing Research: The Researchers' View*. MR-1698-HEFCE. UK: HEFCE and RAND.
- Wouters, P., Thelwall, M., Kousha, K., Waltman, L., de Rijcke, S., Rushforth, A., and Franssen, T. (2015) *The Metric Tide: Literature Review. Supplementary Report I to the Independent Review of the Role of Metrics in Research Assessment and Management*. UK: HEFCE.
- Wunsch-Vincent, S. (2012) 'Accounting for Science-Industry Collaboration in Innovation: Existing Metrics and Related Challenges', in WIPO, *Global Innovation Index 2012: Stronger Innovation Linkages for Global Growth*, pp. 97–107. <https://www.wipo.int/edocs/pubdocs/en/wipo_pub_gii_2012-chapter4.pdf> accessed 28 Mar 2019.