ARTICLE

Research governance and the future(s) of research assessment

Alis Oancea[1]

ABSTRACT   This paper explores recent public debates around research assessment and its future as part of a dynamic landscape of governance discourses and practices, and organisational, professional and disciplinary cultures. Drawing reflectively on data from RAE 2001, RAE 2008 and REF 2014 (reported elsewhere), the paper highlights how recent debates around research assessment echo longer-term changes in research governance. The following changes, and several critiques of their implications, are discussed: shifts in the principles for governing research and the rise of multi-purpose assessment; the spread of performance-based funding and external accountability for research; the use of metrics and indicators in research assessment; the boundary work taking place in defining and classifying units or fields for assessment; the emphasis on research impact as a component of research value; organisational recalibration across the sector; and the specialisation of blended professional practice. These changes are underpinned by persistent tensions around accountability; evaluation; measurement; demarcation; legitimation; agency; and identity in research. Overall, such trends and the discursive shifts that made them possible have challenged established principles of funding and governance and have pushed assessment technologies into a pivot position in the political dynamics of renegotiating the relationships between universities and the state. Jointly, the directions of travel identified in this paper describe a widespread and persistent regime of research governance and policy that has become embedded in institutional and individual practices.

[1] University of Oxford, Oxford, UK. Correspondence and requests for materials should be addressed to A.O. (email: alis.oancea@education.ox.ac.uk)

Introduction

In 2021, sector-wide performance-based research funding in UK academia—and arguably worldwide—will be turning 35, an anniversary also to be marked by the next iteration of the national exercise for the assessment of research performance on higher education. The Research Selectivity Exercise (RSE), conducted in 1986, was the first full-scale national exercise that aimed to base funding decisions on a wide-ranging assessment of the quality of research carried out in university departments. The RSE was criticised at the time on almost every aspect, and many of these critiques led to changes in the design and procedures of its descendant—the Research Assessment Exercise (RAE) in 1989. This pattern of opposition, critique, consultation and amendments is recognisable across all cycles of the exercise, from the RAEs (1989, 1992, 1996, 2001, 2008) to the current Research Excellence Framework (2014, 2021). It is also seen in the striking dynamic of assessment hyperactivity and consultation fatigue that seems to keep academics and administrators too busy to act when yet another spinning plate may be added to their daily jobs. What is also striking is the persistence of debate around the key elements of the exercise, including: the criteria for quality; the indicators of performance and assessment procedures, such as the balance between peer review and metrics; the consultative mechanisms feeding into the design of each round; the treatment of different disciplines and of interdisciplinary work; the extent to which the procedures are sufficiently transparent at all levels; and the impacts on different institutions, fields, modes of research, and categories of staff. The arguments summarised in papers such as Phillimore (1989) and Bence and Oppenheim (2002) are as alive now and they were then, despite much research having been conducted since on the workings and outcomes of successive exercises. According to a great swathe of the literature, the 'dragon of evaluation' (Minogue, 1986)—or the 'Frankenstein' of research assessment (Attwood, 2010)—seems to have grown no more and no less fierce over the years.

How is it, then, that (collective) learning from the very public debates and direct realities of three decades of national exercises is yet to enable academics, policy makers, administrators, 'users' and investors in research to reach agreement on ways to address satisfactorily these recurrent issues?

Part of the answer must be down to the politics and micro-politics of research, higher education, and the 'knowledge economy'. But part of the answer may also be that some of these problems are intractable: not in the sense of strategic stand-offs between the parties concerned, but in the more fundamental sense of the philosophical and sociological tensions that underpin the vocabulary and procedures of measuring, rewarding and influencing research 'performance' or 'quality'.

This paper explores some of these tensions, grouped under seven headings and framed as persistent 'problems' in research assessment as a technology of governance in the shape currently practiced in the UK and to a great extent elsewhere: the accountability; evaluation; measurement; demarcation; legitimation; agency; and identity problems. These problems are philosophical, as well as empirical. Their value as analytic devices derives from the fact that they offer a frame for questioning current and emergent practices and the incentives arising from them. The grounds for this selection are both *a priori*, as themes identified through conceptual and theoretical inquiry into the notions of research quality, performance and value/ evaluation; and *a posteriori*, as categories developed from empirical studies of RAE and REF submission data and of interview and survey data spanning four national exercises: RAE 1996, 2001 and 2008, and REF 2014. While the theoretical arguments and the findings of these studies are reported elsewhere (e.g., Oancea, 2008, 2010, 2014; Oancea et al., 2018), this paper attempts to draw them together

reflectively in an exploration of ongoing trends and discursive threads underpinning recent public debates around research assessment and its future in the UK and beyond.

The accountability problem and the rise of multi-purpose assessment

The past few decades of research policy have seen the ascension of principles, such as formal accountability, marketization, and competition, in the governing of research at international, national and organisational levels. A symptom of this move is the growing reliance on performance-driven assessment technologies not only to inform public investment in research, but also to steer research activity itself towards aims such as global competitiveness and measurable contribution to the 'knowledge economy' (BIS, 2016). This regime of governmentality re-describes the place of research in the world in terms of solutions to externally-defined, global challenges and priorities. In creating these solutions, research is expected to be both academically significant and practically (including politically and economically) astute—qualities which, it is further expected, may be proxied by an ever increasing range of measures and indicators of research quality and impact. I suggest that this project has profound ethical and epistemological implications for the textured, overlapping sets of practices that it attempts to frame and shape. For example, an excessive focus on technical measures of research performance within institutions may influence researchers' perceived freedoms to enact epistemic virtues such as integrity, openness, modesty, circumspection or criticality (see Kerr, 1993 and Battaly, 2013), as well as potentially corralling the moral sense of academic responsibility into performative compliance with managerial and other role responsibilities. Indicators and metrics are not ethically and epistemologically neutral, but the very processes of their creation, use, rejection and renewal may marginalise and displace parts of the research community with lower access to resources and academic capital (Sugimoto and Larivière, 2018).

The changes, over the years, in the explicit purposes of the UK RAE/REF (as stated in the exercises' official guidance documents) illustrate a shift in emphasis. For example, the official guidance for submissions to RAE 2001 stated that the purposes of the exercise were 'to produce ratings of research quality which [would] be used by the higher education funding bodies in determining the main grant for research to the institutions they fund', and to 'inform policy development' (RAE, 1999); the exercise thus was intended to influence public funding bodies and governmental policy bodies. By 2014, this statement had morphed into a range of purposes, including 'to inform the selective allocation of [funding councils'] grant for research'; to 'provide (…) accountability for public investment in research and produce (…) evidence of the benefits of this investment'; and to 'provide benchmarking information and establish reputational yardsticks, for use within the higher education (HE) sector and for public information' (REF, 2011a, 2011b). The legitimate reach of the exercise was thus extended into organisational governance, while its reputational impact was explicitly put on a par with its financial outcomes. Most recently, the Stern consultation about the future of the REF saw the exercise mainly as a tool for the allocation of public funds for research, but accepted that other purposes were also relevant, such as informing institutional strategy and supporting governmental and funding bodies in 'driving research excellence and productivity' (BEIS, 2016a, b).

A key linguistic change between pre-2014 and post-2014 exercises is the explicit mention of accountability as a key purpose of the REF. No mention of accountability was made in the 2001 statement of purpose; by 2008, accountability had made an

appearance in the description of the principles underpinning the conduct of the exercise; but in the 2014 exercise, it became one of its three core purposes: 'The assessment provides accountability for public investment in research and produces evidence of the benefits of this investment.' This is not 'just' semantics. Strong critiques of the RAE and the REF as assessment technologies have revolved around the particular notion of accountability that is assumed to be at their heart, sometimes summed up as 'competitive accountability' (Watermeyer, 2019) or 'performative accountability' (Oancea, 2008), and contrasted to forms of accountability deemed to be better attuned to the values and generative energies of research and researchers. To simplify, two conceptual constellations seem to epitomise this tension: on the one hand, accountability is a formal, mandatory mechanism that is largely vertical (hierarchical) and adversarial, and revolves around (bureaucratic) surveillance, answerability and enforceability. On the other hand, it is conceived as a formative practice, which is horizontal and voluntary, and emphasizes democratic dialogue, communal and collaborative practice, and professional responsibility. Overall, critics and supporters of the REF tend to find affinities with the language associated with one or the other of these contrasting constellations; but note that these concepts do not form a dichotomous choice, but rather create a space populated by a wide range of hybrids, and hence are subject to continued debate. It is this inherent ambiguity that I refer to as the accountability problem.

The role played by the research assessment in the project of governance sketched above is expressed through a versatile balance of powerful practices, including new rituals and routines that affect academic life: 'mock' REFs and 'dry-runs', user panels for internal allocation of funds at HEI level, REF strategy groups and project boards, benchmarking, and so on. As these practices have become more common, the discourses that legitimise them have also become, in turn, less likely to be questioned. While different notions of accountability may be used by those designing and interpreting the purposes of research assessment, in applying the guidance to the exercise, academics may also implicitly shift towards more formalised notions of accountability. This is a soft and pervasive cultural change, working through ambivalent technologies of discipline and self-discipline (in Foucault's sense of the word, as argued in Oancea, 2014), including techniques of cooption and hospitality—for example, via consultations, award ceremonies, nominations and representation on committees and boards, expert panel surveys, consultancies, and stakeholder events. As a result, performance-based funding, the selective distribution of resources and of research capacity, and institutionalised accountability for academic and non-academic impact have become conditional of professional autonomy and self-regulation in higher education.

The emerging sense of consensus around the legitimacy of research assessment as a key mechanism for research funding allocation, accountability, and steering is, however, at best a grudging consensus—more of a truce than a concert. Looking towards the future of research assessment, it continues to be important to question this truce and constantly re-assess the principles of governance underpinning it, including the tensions surrounding different notions of accountability and the impact of policy-driven definitions of research and research quality on the dual support mechanism.

The evaluation problem and performance-based research funding

The past two decades have seen an unprecedented spread of performance-based research funding across the world. Although many countries now use performance agreements, which are based on expectations of future performance (including, in some cases, research performance) and provision for reporting it, there is also widespread use of ex-post performance funding systems, which base grant allocations on assessments of past performance (see Hicks, 2012, Jonkers and Zacharewicz, 2016). Among the latter, the UK large-scale system has been highly influential.

The REF and RAE's international appeal as models of performance-based research assessment for funding purposes arises partly from their system-wide scale and their long history, together with a halo effect from the success in international rankings of the UK research system itself. There are also elements of the design of these exercises that help explain their durability and influence. In particular, there have been repeated expressions of confidence in the quality and fairness of the expert review at the core of the exercises, appreciation of the procedural transparency of the assessment, support for the profile-based aggregate focus (as opposed to individual single ratings), and valuing of the perceived contribution of the exercises to legitimising research as part of academic practice in different types of higher education institutions (HEIs) and disciplines, as well as to increased mutual understanding between the stakeholders involved (see e.g., Coryn, 2007; Hill, 2016).

However, vast swathes of the literature, including some of the literature referenced above, also bring out disadvantages and undesirable consequences of the RAE and REF in particular, and of performance-based research funding systems in general; for example, in terms of the administrative burden involved in running the national exercises, or of the deficit model of academic practice arguably underpinning performance-based steering more generally. This literature points out that the demands set by these exercises through their guidance documents and structures of government and accountability, as they filter through risk-averse layers of organisational management, may generate negative impacts on organisational cultures, on diversity in research and higher education, on the balance between teaching and research, and on individual staff morale and careers (see the diverse positive and negative perspectives reviewed in Oancea, 2010, 2014).

A particularly powerful objection to assessment models underpinning performance-based funding is that they may stimulate 'gaming' (Lucas, 2006) by perversely incentivising a focus on compliance rather than quality, and reliance on agile gamesmanship (Baird and Elliott, 2018) rather than on in-depth experience in developing generative research environments. For example, Lucas (2006) draws on Bourdieu's (1988) idea of academic 'capital' and on Slaughter and Leslie's (1997) analysis of 'academic capitalism' to argue that the 'status and positioning afforded by success in the (RAE) game' have become the *raison d'être* for research in universities, thus cutting to the core values of academic life. Miller (2001, p. 392) argued that 'calculative practices' were 'intrinsic to and constitutive of' the social relations between agents and institutions shaped by technologies of 'governing by numbers', such as costing, standardisation, benchmarking, and performance measurement. Sidhu (2008) draws on this idea to note the seductive power of savvy compliance with audit technologies like the RAE, and the insidious ways in which the calculative practices incentivized by them conspire to re-mould individual and organisational academic identities. Building on Power's (1997) analysis of audits as 'rationalised rituals of inspection' (p. 96), Strathern (2000, pp. 313–314) argues that audit technologies premised on the assumption of measurable, visible performance, such as the RAE, prioritise transparency and 'verification' over the 'real' workings of an institution and its research creativity, thus contributing to 'a leaking away of trust' in expert systems. Overly complex formal accountability 'juggernauts', of which the RAE and REF are seen as examples, may

'create perverse incentives' in the name of transparency and as a result 'are often a source rather than a remedy for mistrust' (O'Neill, 2013: 10, 12; see also Pirrie et al., 2010), thus potentially contributing to the rise of anti-expertise sentiments in public life (Nichols, 2017).

Such criticisms may reflect the fact that the high stakes involved in the assessment of research for funding allocation exacerbate a heuristic tension that is at the heart of any evaluation as practical judgement (De Munck and Zimmermann, 2015): that between *valuing* something (e.g., instances of original and rigorous research); and the courses of action to be taken as a result of a specific, situated *evaluation* process in pursuit of purposes that transcend it (e.g., increased competitiveness or productivity, different kinds of institutional and individual behaviours etc.). In Dewey's (1939) terms, this points to the tension between 'prizing' ('holding dear' or esteeming) and 'appraising' (assigning comparative value relative to other objects in the same category). The former may engender commitment; the latter, compliance. This tension may also be expressed through differing takes on the forms and sources of knowledge and experience required in order to conduct the evaluation itself.

Table 1 illustrates how in the past two decades the assessment of research has not only become more specialised (as indicated by the range of methods and measures now part of formal assessment mechanisms), but also more stratified. Different actors and approaches have clustered around different levels of interest and organisation: from sub-organisational (such as grant proposal evaluations, or staff performance appraisals) to supra-organisational levels (such as sector-wide assessment exercises). The notion of expertise engaged across these strata is not homogenous, but rather split between in-depth topical and methodological understanding of the field of research being assessed and/or of the areas of application relevant to it (i.e., *substantive* expertise, akin to Collins and Evans', 2002, 'contributory expertise') and detailed technical knowledge of the systems of rules, mechanisms, and formalised expectations involved in performing the assessment itself (i.e., *procedural*

expertise, which is distinct from the 'interactional' and 'referred' expertise noted by Collins and Evans, as it falls in a different area of technical expertise, bounded by the structures and norms of the exercise itself). Note, for example, that most job advertisements for REF-related appointments (such as REF directors, managers, officers etc.) in universities include no or little mention of topical or methodological expertise in a particular field or cluster of fields, but expect instead a clear track record of procedural expertise pertaining to the specific details of running the REF. The opposite seems to be true of adverts for most academic positions (including leadership positions) without specific contractual REF responsibilities.

Both forms of expertise are always present in actual assessment (I make no claims about ideal evaluation situations), but in the current governance landscape the balance between the two shifts across different assessment contexts. Arguably, as suggested by the two arrows on the side of Table 1, the closer an instance of research assessment is to individual research projects, 'outputs' and researchers, the stronger its dependence may be on the depth and quality of substantive expertise; and the closer an instance of research assessment is to the other end of the spectrum, the more dependent its actual conduct may be on procedural expertise. Testing these hypotheses may help understand why assessments based exclusively in substantive expertise are often accused of conservatism or self-serving bias; while those heavily dependent on procedural expertise are accused of misinterpretation, intellectual co-option and dulling of critical scrutiny of the assessment itself.

With increased pressures on limited resources from a growing number of organisations (higher education, non-profit institutes, think tanks, for-profit organisations) come incentives to tighten such assessments even further. Research assessment thus balances the policy appetite for rational allocation of resources (which in the recent decades has been interpreted as selectivity and concentration based on performance) with the academic orientation towards intellectually defensible allocation of research prestige (which customarily translates into the outcomes of various forms of peer review).

Table 1 Stratified assessments and forms of expertise (adapted from Oancea, 2009)

Strata of research assessment	Scope of assessment	Purposes	Stakeholders	Methods and measures	
Supra- organisational	International, national, multidisciplinary and disciplinary	Policy and strategic decisions; public resource allocation; political debate; construction of field identity and status; system performance; performance-based funding	International organisations; national government; funding bodies; national strategic bodies; interest groups and sector representative bodies	Performance-based research funding systems; system ratings; economic indicators; bibliometrics; cultural indicators; impact assessments; expert descriptions; scenarios; peer review; Delphi panels; consultation; public debate	Procedural expertise
Organisational	Research organisations; funding schemes; publishing investments; research units; large-scale programmes and partnerships	Benchmarking; accreditation; social accountability; organisational mission and strategy; reputation/ image management; allocation of funds within organisations; human resources management; capacity building	Funding bodies; quality assurance and audit bodies; research organisations; professional/ external evaluators; publishers; industry/ user bodies; third sector; professional associations; media	Formal evaluations; ratings; peer review; bibliometrics; economic metrics; cultural indicators; inter/national standards; impact, engagement and use studies; benchmarking; quality management; consensus conferences; network studies; case studies	
Sub-organisational and para-organisational	Teams; individuals; projects; outputs; outcomes	Access to funds; publication; personnel decisions and workload allocations; awards and recognition; training and development; research decisions (substantive, methodological, practical); decisions on: research synthesis, dissemination, brokerage, KE, PER, open scholarship etc.	Peers; human resources departments; managers; grant awarding bodies; editors and referees; users and partners; brokers; professional associations; etc	Peer review; systematic review; network maps; case studies; public debates; open appraisal; bibliometric counts; alternative metrics etc	Substantive expertise

The measurement problem and the use of metrics and indicators

The third persistent problem identified in this paper is the 'measurement problem', or the problem of whether it is possible to avoid prioritising the metric (or, in semiotic terms, the signifier) over the actual quality (or the signified) in evaluating research (Allan Hanson, 2000, p. 68). In Baudrillard's (1994, p. 2) words, this is a situation where the 'map' produced for evaluation purposes may no longer be a representation or re-imagining of a 'territory' that it purports to describe, but would instead 'precede' it: the 'territory' becomes purely operational, being 'produced from miniaturised units, from matrices, memory banks and command models'—or, one may add, from objects such as units of data, metrics, templates and forms, and codes of practice.

Given the diversity of methods and of measures illustrated in Table 1, the meaning of 'doing well' in research assessment is not straightforward. For example, the official profiles or scores awarded in national assessment exercises are mediated by internal governance factors but also by external referents, and in particular by the position of a research unit relative to any number of comparators. A high score for a research unit in the UK REF does not automatically translate into internal recognition, allocation of resources, or strategic commitment to specific research values within the host HE institution. The result is challenge and flux, with numerous proposals for new metrics vying for primacy to legitimise potential shifts in hierarchies (though note that many of these proposals do not start from a conceptualisation of research value or performance and a theory of how it may be measured, but rather from observing or constructing an aspect of research that is amenable to counting).

Table 2 illustrates how the performative vocabulary that has grown around research metrics varies by scope, from the individual to the field level (and beyond–though the national and international levels are not included in the table); and by level of aggregation, from specific measures of research performance (micro-metrics) to global assessments of research success (macro-indicators). None of the blue boxes in the table is a full list of key metrics and indicators in current use; rather, they offer examples of some of the references to metrics and indicators that I have heard mentioned in the interviews, surveys and workshops I conducted over the past decade. This collection illustrates how what is often called 'metrics' in everyday institutional language may in fact pertain to a range of different categories and may relate only loosely, if at all, to a theory of measurement.

Micro-metrics of research are what is more commonly meant by the term 'metric'. They are measurements of the degree to which research inputs, outputs or outcomes display a particular characteristic. They are usually concrete, quantifiable, time-defined, and narrow in scope. Often, they are co-opted by organisations to function as micro-indicators of performance, in which case their legitimacy is inferred from meso-indicators, macro-indicators, and meta-indicators in an attempt to compensate for their own inherent lack of contextual information and normative self-awareness. As snapshots of a particular moment in time, such functional micro-indicators are of limited direct use in summative judgements of research, despite the bewitchingly normative terminology surrounding their use, such as 'success rate' or 'grant value'. Their transient nature means that they are often the object of constant institutional monitoring over time, despite the doubtful meaningfulness of the resulting reports of quarterly and annual figures, and the seriously damaging implications of their misunderstanding or misuse.

Meso-metrics and indicators are also based on measurable quantities, usually through cumulative measurements of single micro-metrics over time, and with variable degrees of validity and reliability. Meso- indicators play a dual role when used in evaluations: first, they may be drafted in to function as targets for micro-performance (see, for example, the use of publication productivity indicators in academic review and promotion procedures); and second, they may be extrapolated to signal, separately or combined, aspects of performance against macro-criteria—see, for example, the intended use by eleven sub-panels in REF 2021 citation data as a 'potential'(Panel A), 'part(ial)' (Panel B) or 'supplementary' (sub-panel 16 in Panel C) 'indicator of academic

Table 2 The vocabulary of research performance: some examples

	Micro	Meso	Macro	Meta
Field/ discipline	Number of journals Average costs Total staffing figures Total income	Publication volume per (unit) Impact factors Normalised citations Equity measures	Capacity Vitality Diversity Impact	'4-ness' 'Impactfulness'
HEI/ unit	Success rates (funding) Collaborations per unit Income per FTE Downloads and access OA compliance rates	Total research income PhD completions Publication intensity Web analytics Number of awards	Quality/ Rigor/ Originality Significance/ Reach Vitality/ Sustainability	'Grade Point Average' 'Research intensity' 'Research power'
Individual	Grants: number, value, success rate 'Current grant value' Number of publications in press	h-index, i10 index Popularity metrics Altmetrics Volume of peer reviewed publications	Productivity Reputation Leadership	'REF-ability' '3-4 star-ness' '4x4-ness'

significance' (REF, 2018b, pp. 59–60). The assumption under-pinning the latter phrase (which is part of the generic guidance but is toned down in the panel criteria) is that of a (stable) relationship between the frequency of indexed citations and the relationships of esteem in academic communities, and further, between these relationships and shared understandings of quality. Citation theorists such as Wouters (2016) warn that calculated indicators, such as those based on aggregated citation counts, are based on decontextualised information, where meaning is strip-ped off and then constituted anew in the move from the reference embedded in the original text to the reference in the bibliographic list, and then again to the indexed citation, which may be sub-sequently recontextualised for evaluation purposes.

Macro-indicators are global, composite criteria, usually defined at national, disciplinary or international level. Their nature, scope and legitimacy may be subject to continued contestation as fields and modes of research develop. As a result, their assessment requires high levels of substantive expertise and trust and is often largely holistic and qualitative, though it may also be informed by the refinement and integration of collections of meso-metrics and indicators.

Finally, meta-descriptors are artifacts of the assessment exer-cise itself and of the high reputational stakes it raises. They are either post-factum calculations in order to create various league tables out of the results of the RAE/REF (e.g., 'Grade Point Average'), or normative terms used in internal management talk as shorthand for predicted performance in formal assessments (e.g., the "REF-ability' of publications and of examples of impact, or the '4 by 4'-ness of individual researchers, i.e., researchers with four potentially 4* publications at a particular moment in time—usually a REF dry-run or a recruitment or retention decision). As Keane (2003, p. 413) notes, 'signs give rise to new signs, in an unending process of signification'; the temptation is great to ignore the 'variable symbolic significance' of these new signs and treat them as 'quasi-objective indicators of quality, impact and esteem' (Cronin, 2000, p. 450). Many of these terms have entered everyday language and material practices in higher education, administrative organisations, and in the media and social media, often with damaging consequences for research cultures and individual morale. These performative byproducts of assessment continue to thrive in management vernacular inside and outside the HE system, despite growing expressions of organisational commitment to responsible uses of metrics and/or indicators in response to exhortations such as the San Francisco Declaration on Research Assessment in the US–https://sfdora.org/, the Leiden Manifesto for Research Metrics in continental Europe–http://www.leidenmanifesto.org/, or the UK Forum for Responsible Research Metrics–http://www.universitiesuk.ac.uk/policy-and-analysis/Pages/forum-for-responsible-research-metrics.aspx.

The use of metrics in the assessment of research has been subject to heated debate (Rijcke et al., 2016). Much of the debate has been about the technical credibility and the fitness for pur-pose of metrics in research assessment, and revolves, largely, around terminology and method (see Andras, 2011, for a sum-mary). There is something seductive about fine-grained technical arguments about the robustness, accuracy, standardisation, nor-malisation, validity, and reliability of particular quantitative measures of individual and aggregated research performance, however they may be defined; as well as about arguments on the quality, integration/ interoperability, openness and cost-effectiveness of systems and procedures for calibrating, record-ing and organising them. Literature also explores how, in the recent climate for research assessment, metrics and the organi-sational world they purport to measure may be mutually con-stitutive. Kelly and Burrows (2011, p. 130) label this process the 'performative metricisation' of academic practice, whereby

technologies such as the use of metrics 'recursively defin[e] the practices and subjects of university life'. With a nod to Dickens' *Hard Times*, Donovan (2009) describes excessive reliance on metrics over expertise and interpretation as 'Gradgrinding' research activity: over-simplifying the scope and aspirations of research though faith in the objectivity of 'facts' and the effec-tiveness of regulation.

A HEFCE-commissioned review of metrics (Wilsdon, 2016) attempted to chart a middle course between supporters and critics of the use of metrics in research[1]. It found that, particularly in the assessment of research impact and output originality and robustness, current metrics were neither robust, nor a like-for-like replacement for peer review. On this basis, it argued against the wholesale use of metrics for funding, accountability, personnel, strategic and benchmarking purposes, and took a measured view on the benefits of an increased use of metrics to support peer review in the next REF. Instead, the review group recommended the responsible yet restrained use of indicators of aspects of research input, significance and environment, the design, use and interpretation of which must be contextualised to institutional and (inter)disciplinary characteristics and needs, as well as to diverse purposes, levels and scales of assessment. When specific information becomes an indicator of particular 'qualities' of research, it takes on both a reference and a purpose; in other words, it becomes relational and contextual. Acknowledgment of these inherent attributes of indicators ought, in the steering group's view, to stimulate reflexivity, deliberation, a sense of humility, and transparency in key actors' (government, HEI-higher education institutions'- leaders, managers and adminis-trators, funders, publishers, service and infrastructure providers, researchers) use of metrics for the support of scholarly, institu-tional and career diversity in research.

Although the recommendations made by this review and by other initiatives for the responsible use of metrics or indicators are worth heeding, the emphasis on the transparent use and distributed understanding of metrics, particularly if they are not coupled with devolved decision-making and bottom-up influence, may lead to protracted and widespread investment by institutions in refining metrics for top-down capture and quantification of increasingly detailed information. This investment can in itself build commitment and thus become an incentive for wider use of metrics in academia, paving the way for more data-driven gov-ernance in the future. Along the way, the nuances attached to the concept of 'indicators', favoured by the HEFCE review and other initiatives, may become blurred (see Wouters, 2016), and orga-nisational practice may gravitate towards more straightforward reporting of quantitative metrics.

This soft and pervasive change tensions academic identities and their political agency. Academic 'metrics-natives', whose formative years as academics coincided with the rise of perfor-mance monitoring and performance-based funding in research, are pressured (for example, through recruitment and promotion expectations) to assimilate it to their academic habitus from the start of their careers, alongside the outputs-impact-environment and rigour-significance-originality triads of the RAE/REF. Non-natives (either by length of career or by geography) are expected to update and adapt their academic selves, often as a precondition of performing strategic and management roles in their institu-tions. Some embrace metrics, hoping to make assessment less onerous and more equitable, and to make data about and from research more open. Others oppose them as a threat to quality, diversity and professional judgement, and see their use as out of tune with academic norms of scholarly argumentation, criticality and intellectual integrity. Some go with the tide, while acknowl-edging that they felt pressured to 'play safe' for research assess-ment in REF 2014 by sticking to the more easily measurable and

demonstrable (see interview data reported in Oancea et al., 2018), rather than making wider claims for, for example, discursive or cultural contributions from research. Many exercise domesticated resistance while part of the performance management system, and relieved disdain when they no longer need comply.

And so the use of metrics, like that of other assessment technologies, is beset by tensions about what individuals and institutions are trying to get at, how they go about it, from which structural and discursive positions, and to what purpose and effect. That is because, when integrated in particular performance regimes, metrics and indicators become multiply ambivalent technologies (a term I explained in detail in Oancea, 2014). These rankings, criteria, metrics and indicators are not meaningful on their own, but are ascribed meaning as part of wider narratives, institutional practices and flows of power at different levels and for different entities and purposes. They play out in distinctive ways in governance processes. The issue is not just technical—which metrics and indicators to throw in the basket and how to fine-tune them—but also substantive and normative: what do they mean, to whom and in what structural conditions, why are they seen to matter, whose view takes precedence, and for what purposes and in what context are they mobilised? The reason behind this ambivalence of metrics, however responsibly used, is that they are inevitably drafted into an ongoing renegotiation of the principles underpinning the relationships between universities and the state, mediated through public funding arrangements. Excessive focus on technical issues of measurement may distract from more fundamental debates around the ways in which highly formalised, complex performance assessment systems may affect these principles.

The demarcation problem and boundary work in research assessment

Sector-wide machineries for research assessment, like the UK REF, are engaged in bounding and curating judgements about epistemic objects such as research outputs and bodies of research. They rely heavily on peer review by academics with substantive expertise in the specific fields or subfields covered by each component unit of the exercise (reflected in the definition of panels and sub-panels), sometimes supplemented by field-relevant metrics and indicators. The mechanics of the REF (see Derrick, 2018) tie both the peer review and the use of metrics into definitions and classifications of research fields. The exercise entails decisions about what substantive and methodological content is to be assessed, what expertise that assessment requires, what the yardsticks and comparators ought to be, and what is to be passed on to other experts as not clearly within the remit of a sub-panel. Inevitably, the history of these decisions becomes constitutive of how research is valued, selected and prioritised by HEIs in preparation for submission: the boundaries determined by the assessment machinery through mechanisms of definition and classification are ultimately interpreted, internalised and policed through selection decisions made at research unit level. I have titled this problem the 'demarcation problem' as a nod towards the long-standing debates in the philosophy of science about the grounds for distinguishing between science and pseudoscience; but my interest here is to highlight the sociological rather than epistemological implications of classificatory practices concerning epistemic objects.

Examples of such boundary work occurring early in any given REF cycle include the funding councils' initial decisions about what Units of Assessment are to be evaluated. This initial classificatory work often engages academic voices through a consultation process about what boundaries had or had not worked in a previous round; see, for example the responses to bringing

together Geography, Environmental Studies and Archaeology in a single REF 2014 sub-panel, and their separation again in 2021. Further, individual 'areas of expertise' are taken into account in the appointment of panel and sub-panel members; a pre-definitional process that no longer directly engages academic voices (except for the appointment of the panel chair), as HEIs are not able to contribute directly to the nomination process and instead this tasks falls to learned societies, professional bodies and other agents. The initial selection of these areas of expertise pre-dates the sub-panel's work on scoping the field, and has gained more weight in the preparation for REF 2021, as only a small subset of the final sub-panel members will be contributing to defining the scope, criteria and ways of working of the sub-panels.

Sub-panels' work on defining the scope of the unit of assessment is probably the clearest example of definitional boundary work in the REF, as the succession of RAEs/REFs has produced definitions of fields and sub-fields of research enshrined in official guidance documents. Most definitions include lists of sub-fields and approaches that are 'included' in a particular sub-panel's remit (REF, 2012). The operation of the actual evaluation, in particular the decisions about allocation of outputs, cross-referrals between panels, the moderation and calibration of assessment, the addition of further sub-panel members or assessors in the assessment phase only, and, in the forthcoming REF 2021, the output 'flagging' mechanism ('interdisciplinary identifier') and the input from interdisciplinary panel and sub-panel advisors (REF, 2018a, 2018b), pulls this boundary work in different directions, through competing pressures both to rigidify and to loosen disciplinary boundaries. Although interdisciplinary or multidisciplinary research have always been eligible for submission, there is some evidence that a broadly interdisciplinary submission to a REF 2014 sub-panel that had already reached a consensus view on what its field of assessment encompassed may have been seen as high-risk by strategic institutional leaders (Technopolis/ SPRU, 2016; BEIS, 2016a). The outcomes of the more detailed procedures for the assessment of interdisciplinary research in REF 2021 remain to be seen.

The use of metrics and indicators to inform peer review in some panels is another space for boundary work prior to and during the evaluation. As indicated by the bibliometrics pilot prior to the REF 2014, citation indicators are not seen as technically suitable unless they are field-normalised and also contextualised in relation to protected characteristics. This raises the question of what counts as a 'field' of research. The definition of bibliometric fields in the REF is usually constrained by technical decisions already made by database providers in creating the data infrastructure that makes citation indicators possible in the first place—pre-defined subject categories, research areas and/or research fields are used by commercial databases of research publications to classify journals and papers (particularly in the case of papers published in journals deemed multidisciplinary). Often these fields correspond to disciplines and sub-disciplines, perhaps in line with subject classifications in other library and information data systems, other times they cluster information about the citation network of a paper. 'Multidisciplinary' research may also form a category in its own right (for example, to be used for generalist journals that cover a range of sciences), but in the bibliometrics pilot for the 2014 REF papers published in such journals that had not already been reclassified by the database provider were reassigned to 'more appropriate categories' (p. 13). Such reassignment fits the above definition of boundary work.

At HEI level, preparation for REF submission also involves boundary work. Within HEIs, the allocation of staff to units of assessment entails boundary negotiations relative to both the definitions of units of assessment and eligibility criteria in the REF guidance, and internal clusterings of areas of research and

teaching subjects. Moreover, in some disciplines the expectation to submit impact case studies has also generated further boundary work. For example, the impact case study form separates research from its impact, which may pose particular challenges in art-related fields where research and impact may be organically embedded in creative practice or practice-as-research (see examples in Oancea et al., 2018; Adams and McDougall, 2015).

REF-related boundary work continues after the end of the evaluation phase, too. As a final example, the funding formulae underpinning the application of resources post-REF use multipliers which are based on the allocation of disciplines to three different cost bands, with clinical and laboratory subjects, including psychology, classified in the top cost band and most social sciences and humanities in the lowest one. It is unclear whether the funding formulae reward the participation of humanities and social science units in interdisciplinary work across subjects with different cost weightings.

The outcomes of the boundary work illustrated by these examples become constitutive of everyday understandings and strategic priorities in research units across the sector. Paradoxically, as the mechanics of the exercise depend on mechanisms of differentiation such as disciplinary definitions and classifications, they can also lead to a false perception of comparability across units of assessment, with post-REF calculations of aggregate scores and meta-descriptors being used for marketing or for internal allocation of resources in apparently discipline-neutral ways.

The legitimation problem and the rise of impact as a component of research value

The institutional legitimacy of the REF—the extent to which it is accepted as 'authoritative, binding or valid' (Gellner, 1974, p. 24) in underpinning decisions—depends on both the scientific and the political legitimacy of (partly or fully) publicly-resourced research; hence the necessary reliance on peer review in the conduct of the exercise, and on political processes in effecting its outcomes. Under the cumulative discursive weight of successive assessment exercises, research itself, as it is understood in the public space, has been reframed and re-defined, from a focus on research 'understood as original investigation undertaken in order to gain knowledge and understanding' in both RAE 2001 and 2008 (RAE, 1999, and RAE, 2005), to its being 'defined as 'a process of investigation leading to new insights effectively shared" (REF 2011a, 2011b, REF 2012, REF 2017). Arguably, this definitional change opened a place for knowledge sharing, exchange and impact right at the heart of policy understandings of the nature and value of research activity.

The introduction of impact in the assessment framework may thus be seen as a mechanism for indirect legitimation of the regulatory framework itself, through re-arranging the discursive construction of research excellence in ways that are rooted in both scientific and political epistemic communities (see Filippakou, 2017). This shift reflects wider and longer-term policy discourses about the relationships between higher education and industry, the connections between academic and non-academic contexts, the relevance of research to users and the wider role of research in the so-called knowledge and innovation society/economy (as evidenced by a long succession of white papers, reports and other policy documents—see Lockett et al., 2015)—with added strength drawn from discourses rooted in professional cultures about evidence-based or evidence-informed practice (and more recently, policy) in professions such as medicine or education (see Nunan et al., 2017). Impact had also been for a sometime a priority along both arms of the UK 'dual support' system for research—however challenged the principles behind

the system itself might be in the face of ongoing structural, political and financial pressures to align with each other. The Royal Charters of the Research Councils and their strategic frameworks, as they stood until the 2017 HE bill, already drew direct links between good research and social, cultural, health, economic and environmental impacts. The Councils were interested in impact largely prospectively, in terms of plans and potential benefits, but also retrospectively, with ever closer scrutiny and reporting of impact post-award and after the end of award. In some ways, the REF's falling into step with impact in 2011 amplified an agenda that was already pervasive.

For the purposes of the REF 2014, impact was defined as 'an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia' (REF, 2011, 2012) and was assessed by academic and user reviewers on the basis of standard-format case studies and unit-level strategic statements, using the twin criteria of 'reach' (or breadth) and 'significance' (or depth) of impact. The definition (plus some further explication) and the criteria have remained the same for REF 2021. At 20% of a unit's final 'quality profile' in the REF 2014 (25% in REF 2021), impact has become a weighty element of the financial and reputational hierarchies at stake. Its introduction as one of the three domains for the assessment of research in the UK Research Excellence Framework in 2014 has had mixed, but lively, responses (see Chubb and Reed, 2018, for a review of the different positions and Collini, 2012, for a critique).

It may be argued that the addition of impact to system-wide research evaluation is a further form of boundary work, potentially leading to epistemological problems. After all, an impactful department will be valued more highly and receive higher reward than another department with exactly the same ratings as the first one for outputs and environment (Battaly, 2013). According to Battaly (2013), this situation may indicate 'epistemic insensibility', in the sense of implicitly signalling that research with less evident or direct impact may be less valuable to institutions in constructing their narrative of success - and by extension, letting percolate a sense of its being of lower epistemic value, and decreasing the propensity to engage in the appropriate epistemic practices associated with it. While the above is only one of the several strands of criticisms associated with the assessment of research impact, it stands in telling contrast with views of research assessment as epistemically neutral—in other words, as being concerned with the pragmatic allocation and justification of resources only, rather than with apportioning epistemic value.

Beyond the practicalities of the assessment exercise, the emphasis on impact can be seen as both a driver and an outcome of public renegotiation of the values that underpin the case for public investment in research. In recent years, as professing mistrust in expertise, truth, facts or academic rigour became more politically fashionable, impact has grown in discursive importance, although particularly in instrumental guises that may fit a range of normative frames and may be at odds with the wider aims of impact assessment, including those stated in the context of the REF (REF, 2012 and REF, 20018a, 2018b; see McCowan, 2018, for an extension of this argument). Hence the increasing political emphasis on research impact as a component of research value has come in tandem with academic critiques of the instrumentalisation and monetisation of research.

The agency problem and organisational recalibration
Research assessment, be it for performance-based funding such as the REF or in the light of requirements of key research funders, such as the Research Councils, government departments, or the European Commission, has been one of the drivers of the rise of

research and knowledge exchange as part of the institutional mission of HEIs in the UK over the past few decades. Research on the impacts of the RAE/REF suggests that the exercises may have contributed to strengthening research cultures and the volume and quality of research and research communication in many institutions, but that they may have also affected the nexus between teaching and research, in particular in undergraduate provision, and may have increased the likelihood of a more pressurised and unequitable climate in a range of institutions (see e.g., Oancea, 2014).

The high-stakes status (reputationally and financially) of the REF outcomes has had major implications for the everyday work of HEIs, amounting to wide-ranging organisational recalibration across the sector. HEIs have flexed, stretched or contracted to accommodate the ever-evolving definitions of performance. Some of these changes have affected directly the capacity for research in institutions, for example through recruitment drives, changes to the contractual arrangements of staff (leading in some cases to defined separation between the workloads of teaching only and research active staff), or through the inclusion of outputs and, now, of impact among the criteria for the recruitment and promotion of staff, particularly to senior positions. As evidenced by the submissions to the REF (see e.g., Mills, Oancea and Robson, 2017), the workload models in many institutions have been adjusted to make space for impact activity—including 'pathways' to impact such as managing relationships of partnership, knowledge exchange, dissemination, or public engagement with research activities. New senior academic responsibilities have emerged: Impact Champions, Directors and Deans for Impact, Knowledge Exchange Leads, together with further appointments of Professors of Public Understanding of Science (and cognate titles), and so on.

The introduction of impact in the REF 2014 shaped strategic decisions in HEIs to invest differentially in areas of research, to restructure the organisational basis for the provision, validation and sharing of research, and may have contributed to re-directing parts of research activity towards shorter horizons of contribution to political priorities and societal challenges (as documented in Oancea, 2013). It also influenced decisions about the size and shape of the REF submissions themselves. Within the logic of the assessment exercise, in the lead up to 2014 the need to submit around one impact case study per ten FTE 'research active' staff in a unit prompted a lot of effort to generate, corroborate and write viable case studies, but also tactical decisions among units to have more or less inclusive submissions. For example, Kerridge (2015) notes a 'spike' in submissions just under each of the FTE thresholds beyond which a further impact case study would have been required.

The strategic leadership, management and governance of research in universities have also been recalibrated. The environment and impact statements submitted to the REF 2014 (see e.g., those analysed in Mills, Oancea and Robson, 2017) show that institutional managers think strategically about and monitor and scrutinise closely the research activity in their unit/s. Research strategies encompass, for example, incentive structures for research engagement and productivity at different stages of career; specific steering towards unit-level (rather than individually determined) substantive and methodological foci; collective output and publication plans; as well as tactics for attracting external research income. The more recent addition, in view of the REF 2021, of open access requirements has generated an unprecedented level of monitoring of publication cycles, which has been embedded in institutions with much more ease than many other changes, possibly due to the fact that it taps into shared values of fairness, freedom and visibility of research knowledge.

The outcomes of the exercise, actual or anticipated, have also lead to recalibration. Reputational outcomes may open or close possibilities for organisational growth, partnerships, or student recruitment; while financial outcomes may sustain or damage the vitality of established research environments and research capacity (for example, in universities with a history of significant quality-related funding), but they may also be the impetus for (or dampener of) emergent growth.

Overall, both the process and the outcomes of performance-based research funding have tensioned the organisational ethos of institutions, as they internalised their localised interpretations of the funders' requirements. Many institutions have made difficult choices in the light of these interpretations, for example between distributed (but fragmented and slow) and hierarchical (but instrumentally efficient) governance structures; between potentially divisive (but sharp) or more cohesive (but generic) strategic priorities and mission statements; and between transparent (but endlessly redressed) or opaque (but contentious) mechanisms for the management and administration of research and of research funding. This way, what is being measured and monitored and what matters to researchers and their communities have subtly morphed into each other. Both principled resistance and pragmatic compliance may be ultimately co-opted by the strategic institutionalisation of research assessment in organisations faced with the rigours of performance-based allocations of funding and recognition. It may seem easy to dismiss these decisions as defensive routines within organisations. However, as responsibility for these transformations is regularly passed backwards and forwards along structural and political lines (for example, between government agencies, funding bodies, and different layers of institutional management), the problem remains, of who ultimately owns this agenda and who has the agency to introduce or reverse change in organisational practices and trans-organisational networks.

The identity problem and the growth in blended professional practice

In the UK, a clear area of specialisation has formed around the RAE/REF, with many UK higher education institutions, as well as the bodies tasked with allocating their core funding, appointing senior REF directors, project managers or administrators, and other dedicated staff for the different aspects of academic performance recognised in the REF. For example, in response to the addition of impact assessment to the REF in the 2010 pilot and 2011 guidance, most institutions have added RE impact-related tasks to existing roles, including those of directors, deans and other senior research management staff, and have created impact task forces, project boards, and delivery and oversight groups. They have created new roles or reframed existing ones, such as impact officers, KE officers, professional impact writers, case study copy-editors, and public engagement managers (see Manville et al., 2015). They have also employed a large number of casual workers (many of whom are postgraduate students) to collect, input and clean data on impact and on different metrics. While a large proportion of such posts created prior to REF 2014 were temporary and there has been vast restructuring and mobility in these areas since, many were not and have since become established parts of organisational structures, with many impact and assessment professionals appointed during the previous cycle now line-managing new colleagues or entire units. Even in institutions where the pre-2014 appointments had been fixed-term, the model remained inscribed in their REF planning documents, and in many cases it is being revived as preparations get underway for the next exercise.

As a way of tooling the new impact- and research monitoring-related practices, some institutions have bought into the thriving

market of commercial packages for monitoring and recording research and impact activities (or have created their own packages), and have invested in the training and allocation of staff time necessary to operate them. Further investments into the growing 'para-academic' (Macfarlane, 2011) industry associated with performance-based assessment include the buying in of experience in the form of expert advisors and external reviewers for the running of 'mock' REF exercises and the decoding of REF guidelines.

In this context, impact-related staff—with the exception of many of the precarious workers drafted into supporting the basic rungs of running the exercise—have strengthened their professional identity and sense of community over the recent years, perhaps echoing the way in which research management became a fully recognised area of professional HE practice in the past few decades, supported by the stronger voice of professional organisations such as ARMA (Association of Research Managers and Administrators, incorporated in 2006 and the predecessor of which was created in 1991).

In addition, the ongoing arrangements for performance based research funding have stimulated the increasing professionalisation of other specialist 'blended' (Whitchurch, 2009) or 'third space' (Whitchurch, 2012) practice and practitioners, such as industry partnership, entrepreneurship, or commercialisation managers. These 'dedicated appointments spanning professional and academic domains' (Whitchurch, 2009, p. 408) have developed widely in different sectors (including public, commercial and third sector research), across different aspects of institutional missions (beyond research), and in different international contexts. 'Braided' careers that alternate between, or otherwise combine, work in academia and other sectors are becoming more common. Secondments to and from other sectors, and various visiting positions, internships and practitioner or industry fellowships spanning the boundaries between HEIs and other types of organisation, are also used to facilitate the 'brokering' of new research networks and quasi-formal relationships with the potential to generate collaborative research and impact.

Arguably, the growth of such relationships contributes to 'unbundling' (Macfarlane, 2011, Locke, 2014) current constructions of academic identities and careers, and to introducing further (and welcome) differentiation. At the same time, it may lead to miscommunication and territorialism, as the spaces occupied by research professionals and professional researchers get renegotiated; or to new forms of inequality, as the gaps between specialist career tracks and precarious academic work may widen.

What next?

Coupled with wider 'soft' practices in the governance of research, including co-option and hospitality, assessment technologies like those discussed in this paper are Janus-like, operating in a transient equilibrium that is highly sensitive to changes in the domestic and international research economy. I argued before that technologies like the REF are 'multiply ambivalent': they place individual, institutional and trans-institutional forms of participation, responses, and consequences in a versatile balance of overlapping tensions (Oancea, 2014). These ambivalences are not only down to the pragmatic details of how the REF is practiced in everyday institutional contexts, but are also traceable to persistent sociological and philosophical problems and systemic structural issues that underpin high-stakes, large-scale assessments of research performance. In this paper, I highlighted seven such problems, and I connected them with a consideration of the directions of travel in research assessment that I detected in my empirical research on research. Collective learning from the experience of several decades of formalised sector-wide research

assessment in the UK, particularly through full consideration of insights from relevant research on research, may help ground these debates. A large body of empirical theoretical and critical literature has already been developed around research policy and assessment; while government departments and funding councils have also commissioned a range of evaluations and reports, including the Stern review and *Metric Tide* report (BEIS, 2016b, Wilsdon, 2016). This paper has drawn reflectively on findings from past research to make a contribution to this collective learning project.

Overall, the directions of travel identified in this paper and the discursive and political shifts that enabled them have challenged established principles of funding and governance, including the dual support, and have pushed assessment technologies into a pivot position in the political dynamics of renegotiating the relationships between universities and the state. Transformative change of the research governance regime discussed in this paper and of its implications, while possible, would be a major undertaking, which could not rest on simply removing any one particular element of it, but would need to involve changing both the structural conditions that underpin it, and the cultural and normative premises that legitimise it. As far as glimpses into the future go, the UK seems to have placed its bets on performance-based resource allocation and funding-based incentivisation of organisational and individual behaviour; complex and formal accountability systems; and an emphasis on extra-academic definition of research agendas and valuation of their outcomes. In the light of current geo-political changes and regional power re-configurations, these mechanisms are seen as key means for sustaining the capacity for, and quality of, research in the UK. To achieve this goal, however, balanced funding policies and a diverse portfolio of funding opportunities would need to be coupled with a determined stance on enabling healthy governance in the research and higher education system. The pre-conditions for such governance include intellectual freedom in research; structural conditions for insightful, dialogical, equitable and responsible decision-making; support and recognition for a truly diverse and critical academic agora; and commitment to the public funding of diverse modes of higher education research (including research that is critical, theoretical and conceptual, expressive or interpretive, and goes beyond short-term political agendas).

At the same time, a swell of generative energies from across all strata of the research communities is now pushing for active and more radical re-imagining of the organisation of research and research assessment, of its structures and mechanisms, and of its norms and values. Arguments are bubbling up for re-balancing intrinsic and extrinsic interpretations of value, for recognising fully and supporting structurally the epistemic value of diversity, and a richer sense of equity, and for nurturing the symbiotic relationship between freedom and responsibility. These are not escapist, nor 'alternative', voices to be othered or dismissed, but principled movements towards re-claiming the moral and intellectual strengths of academic research. A strong research-on-research base, genuine dialogue and courageous leadership would be necessary in order to re-imagine research assessment as a formative, communicative, epistemically sound and morally defensible enterprise.

Data availability

Data sharing is not applicable to this paper as no new datasets were generated or analysed.

Notes

1 Parts of this section are adapted, with permission, from a piece published originally in *Research Fortnight* (Oancea, 2015).

References

Adams J, McDougall J (2015) Revisiting the evidence: practice submissions to the REF. J Media Pract 16(2):97–107. https://doi.org/10.1080/14682753.2015.1041803

Allan Hanson F (2000) How tests create what they are intended to measure. In Filler A(ed) Assessment: social practice and social product. Routledge Falmer, London and New York

Andras P (2011) Research: metrics, quality, and management implications. Res Eval 20(2):90–106

Attwood R (2010, September 9) 'Frankenstein' assessment is out of control. *Times Higher Education*

Baird JA, Elliott V (2018) Metrics in education—control and corruption. Oxford Review of Education 44(5):533–544. https://doi.org/10.1080/03054985.2018.1504858

Battaly H (2013) Detecting epistemic vice in higher education policy: epistemic insensibility in the Seven Solutions and the REF. J Philos Educ 47(2):263–280

Baudrillard J (1994) Simulacra and simulations. The University of Michigan Press, Ann Arbor

Bence V, Oppenheim C (2002) The evolution of the UK's Research Assessment Exercise: publications, performance and perceptions. J Educ Adm Hist 37 (2):137–155

BIS (2016) Success as a knowledge economy: teaching excellence, social mobility and student choice. White Paper, Cm 925, BIS/16/265. Department for Business, Innovation and Skills, London

BEIS (2016a) Lord Stern's review of the Research Excellence Framework Call for evidence. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/500114/ind-16-1-ref-review-call-for-evidence.pdf

BEIS (2016b) Research Excellence Framework (REF) review: building on success and learning from experience (the Stern review). Department for Business, Energy and Industrial Strategy, London

Bourdieu P (1988) Homo academicus. Polity Press, Cambridge

Chubb J, Reed MS (2018) The politics of research impact: academic perceptions of the implications for research funding, motivation and quality. Br Polit 1–17, https://doi.org/10.1057/s41293-018-0077-9

Collini S (2012) What are universities for? Penguin, Harmondsworth

Collins HM, Evans R (2002) The third wave of science studies: studies of expertise and experience. Social Stud Sci 32(2):235–296. https://doi.org/10.1177/0306312702032002003

Coryn CLS (2007) Evaluation of researchers and their research: Toward making the implicit explicit. Doctoral dissertation, Western Michigan University, Kalamazoo

Cronin B (2000) Semiotics and evaluative bibliometrics. J Doc 56(4):440–453

De Munck J, Zimmermann B (2015) Evaluation as practical judgment. Human Stud 38:113. https://doi.org/10.1007/s10746-014-9325-1

Derrick G (2018) The evaluators' eye: impact assessment and academic peer review. Palgrave Macmillan, London

Dewey J (1939) Theory of valuation. In JA Boydston (Ed.) The Later Works of John Dewey, Vol 13. Southern Illinois University Press, Carbondale, pp. 189–251

Donovan C (2009) Gradgrinding the social sciences: the politics of metrics of political science. Political Stud Rev 7:73–83

Filippakou (2017) The evolution of the quality agenda in higher education: the politics of legitimation. J Educ Adm Hist 49(1):37–52. https://doi.org/10.1080/00220620.2017.1252738

Gellner E (1974) Legitimation of belief. London: Cambridge University Press. Pring, R. & Thomas, G. (Eds) 2004. Evidence-Based Practice in Education. Maidenhead: Open University Press

Hicks D (2012) Performance-based university research funding systems. Research Policy 41(2):251–261. https://doi.org/10.1016/j.respol.2011.09.007

Hill S (2016) Assessing (for) impact: future assessment of the societal impact of research. Pal Comm 2, https://doi.org/10.1057/palcomms.2016.73

Jonkers K, Zacharewicz T (2016) Research performance based funding systems: A comparative assessment. Brussels: European Commission, EUR 27837 EN. https://doi.org/10.2791/659483

Keane W (2003) Semiotics and the social analysis of material things. Lang Commun 23(3):409–425

Kelly A, Burrows R (2011) Measuring the value of sociology? Some notes on performative metricization in the contemporary academy. Sociol Rev 59 (2s):130–150

Kerr C (1993) Higher education cannot escape history: issues for the twenty-first century. State University of New York Press, Albany

Kerridge S (2015, February 11) How thresholds for case studies shaped REF submissions. *Research Fortnight*

Locke W (2014) Shifting academic careers: implications for enhancing professionalism in teaching and supporting learning. Higher Education Academy, London

Lockett A, Wright M, Wild A (2015) The institutionalization of third stream activities in uk higher education: the role of discourse and metrics. Br J Manag 26:78–92

Lucas L (2006) The research game in academic life. Open University, Maidenhead

Macfarlane B (2011) The morphing of academic practice: unbundling and the rise of the paraacademic. High Educ Q 65(1):59–73

Manville C, Morgan Jones M, Frearson M, Castle-Clarke S, Henham ML, Gunashekar S, Grant J (2015) Preparing impact submissions for REF 2014: an evaluation: findings and observations. HEFCE, London

McCowan T (2018) Five perils of the impact agenda in higher education. Lond Rev Educ 16(2):279–295. https://doi.org/10.18546/LRE.16.2.08

Mills D, Oancea A, Robson J (2017) The Capacity and Impact of Education Research in the UK. Report to the Royal Society and British Academy Joint Enquiry on Educational Research. London: RS/BA

Miller P (2001) Governing by numbers: why calculative practices matters. Social Res 68(2):379–396

Minogue K (1986) Political science and the gross intellectual product. Gov Oppos 21:396–405

Nichols T (2017) The death of expertise. Oxford University Press, Oxford

Nunan D, O'Sullivan J, Heneghan C, Pluddemann A, Aronson J, Mahtani K (2017) Ten essential papers for the practice of evidence-based medicine. BMJ Evid-Based Med 22:202–204

Oancea A (2009) Standardisation and versatility in research assessment. In Besley A(ed) Assessing the quality of research in higher education. Sense, Rotterdam

Oancea A (2008) Performative accountability and the UK Research Assessment Exercise. ACCESS: Critical Perspectives on Communication, Cultural & Policy Studies, 27(1/2): 153–173

Oancea A (2010) The Impacts of RAE 2008 on Education Research in UK Higher Education Institutions. Macclesfield: UCET/BERA

Oancea A (2013) Interpretations of research impact in seven disciplines. Eur Educ Res J 12(2):242–250. https://doi.org/10.2304/eerj.2013.12.2.242

Oancea A (2014) Research assessment as governance technology in the United Kingdom: findings from a survey of RAE 2008 impacts. Z Fur Erzieh 17:83–110. https://doi.org/10.1007/s11618-014-0575-5

Oancea A (2015) Metrics debate must be about ethics as well as techniques. *Research Fortnight*

Oancea A, Florez-Petour T, Atkinson J (2018) The ecologies and economy of cultural value from research. Int J Cult Policy 24(1):1–24. https://doi.org/10.1080/10286632.2015.1128418

Oancea A (2016) Challenging the grudging consensus behind the REF. Times Higher Education, 25 March

O'Neill O (2013) Intelligent accountability in education. Oxf Rev Educ 39(1):4–16

Phillimore AJ (1989) University research performance indicators in practice: The University Grants Committee's evaluation of British universities, 1985-86. Res Policy 18:255–271

Pirrie A, Adamson K, Humes W (2010) Flexing academic identities: speaking truth to power. Power Educ 2(1):97–106

Power M (1997) The audit society: rituals of verification. Oxford University Press, Oxford

Sidhu R (2008) Risky custodians of trust: Instruments of quality in higher education. Int Educ J 9(1):59–71

Slaughter S, Leslie L (1997) Academic capitalism: politics, policies and the entrepreneurial university. The John Hopkins University Press, Baltimore

Strathern M (2000) The tyranny of transparency. Br Educ Res J 26(3):309–321

Sugimoto CB, Larivière V (2018) Measuring Research. What everyone needs to know. Oxford University Press, Oxford

Technopolis/SPRU (Science Policy Research Unit, University of Sussex) (2016) Landscape Review of Interdisciplinary Research in the UK. Report to HEFCE and RCUK. London: HEFCE

RAE (1999) Guidance on Submissions RAE 2/99. HEFCE, London

RAE (2005) Guidance on Submissions RAE 03/2005. HEFCE, London

REF (2011a) Decisions on assessing research impact. REF 01.2011. HEFCE, London

REF (2011b) Assessment framework and guidance on submissions. HEFCE, London, REF02.2011, July

REF (2012) Panel criteria and working methods. HEFCE, London, REF01.2012, Jan

REF (2017) REF 2021 Decisions on staff and outputs. HEFCE, London, November

REF (2018a) Draft guidance on submissions (2018/01). Research England, London, 23 July

REF (2018b) Consultation on the panel criteria and working methods (2018/02). Research England, London, 23 July

Rijcke S, de, Wouters PF, Rushforth AD, Franssen TP, Hammarfelt B (2016) Evaluation practices and effects of indicator use—a literature review. Res Eval 25(2):161–169

Watermeyer R (2019) Competitive accountability in academic life: the struggle for social impact and public legitimacy. Edward Elgar, Cheltenham, forthcoming

Whitchurch C (2009) The rise of the blended professional in higher education: a comparison between the UK, Australia and the United States. High Educ 58 (3):407–418

Whitchurch C (2012) Reconstructing Identities in HE: the Rise of the 'Third Space' professionals. Routledge, London

Wilsdon J (2016) The metric tide: independent review of the role of metrics in research assessment and management. SAGE, London, (chair)

Wouters P (2016) Semiotics and citations. In Sugimoto CR(ed) Theories of Informetrics and Scholarly Communication. A Festschrift in honour of Blaise Cronin. De Gruyter, Inc, Berlin, p 72–92

Acknowledgements

Additional information

Competing interests: After the acceptance of this paper, the author became a member of the Research England Advisory Group on the 'Future of Research Assessment' (2019).

The author is REF2021 coordinator for Unit of Assessment 23 at the University of Oxford.

Reprints and permission information is available online at http://www.nature.com/reprints

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.