



How does research evaluation impact educational research? Exploring intended and unintended consequences of research assessment in the United Kingdom, 1986–2014

European Educational Research Journal

2017, Vol. 16(6) 820–842

© The Author(s) 2017

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1474904117730159

journals.sagepub.com/home/eeerj



Marcelo Marques

Institute of Education and Society, University of Luxembourg

Justin JW Powell

Institute of Education and Society, University of Luxembourg

Mike Zapp

Institute of Education and Society, University of Luxembourg

Gert Biesta

Brunel University London, UK

University of Humanistic Studies, the Netherlands

NLA University College, Norway

Abstract

Research evaluation systems in many countries aim to improve the quality of higher education. Among the first of such systems, the UK's Research Assessment Exercise (RAE) dating from 1986 is now the Research Excellence Framework (REF). Highly institutionalised, it transforms research to be more accountable. While numerous studies describe the system's effects at different levels, this longitudinal analysis examines the gradual institutionalisation and (un)intended consequences of the system from 1986 to 2014. First, we analyse historically RAE/REF's rationale, formalisation, standardisation, and transparency, framing it as a strong research evaluation system. Second, we locate the multidisciplinary field of education, analysing the submission behaviour (staff, outputs, funding) of departments of education over time to find decreases in the number of academic staff

Corresponding author:

Marcelo Marques, Institute of Education and Society, University of Luxembourg, Maison des Sciences Humaines, 11, Porte des Sciences, L-4366 Esch-sur-Alzette, Luxembourg.

Email: marcelo.marques@uni.lu

whose research was submitted for peer review assessment; the research article as the preferred publication format; the rise of quantitative analysis; and a high and stable concentration of funding among a small number of departments. Policy instruments invoke varied responses, with such reactivity demonstrated by (1) the increasing submission selectivity in the number of staff whose publications were submitted for peer review as a form of reverse engineering, and (2) the rise of the research article as the preferred output as a self-fulfilling prophecy. The funding concentration demonstrates a largely intended consequence that exacerbates disparities between departments of education. These findings emphasise how research assessment impacts the structural organisation and cognitive development of educational research in the UK.

Keywords

Educational research, research evaluation system, higher education, Research Assessment Exercise, Research Excellence Framework, United Kingdom

Introduction: Institutionalising research evaluation

Contemporary shifts in the governance of higher education and science systems, including research evaluation, have been examined in a growing body of literature (see, e.g., Besley and Peters, 2009; Geuna and Martin, 2003; Hicks, 2012; Martin and Whitley, 2010; Oancea, 2008). Elaborate evaluation systems have been implemented as policy tools to allocate public funding and attempt to safeguard research quality through evaluation. The United Kingdom's Research Assessment Exercise (RAE), rebadged as the Research Excellence Framework (REF) for the 2014 evaluation cycle, is among the most highly institutionalised performance-based research funding systems worldwide. Supranational bodies (European Commission, 2010; OECD 2010) and other countries have been inspired by the REF in developing their own research evaluation systems; it remains an influential model (see Schneider and Bloch 2016; Karlsson, 2017). Over the past several decades, Spain's *Sexenios de investigación* (1989), Portugal's *Research Unit Evaluation* (1996), *Excellence in Research for Australia* (2003), and New Zealand's *Performance-based Research Funding* (2006), to name a few, underscore the global diffusion of formalised research evaluation. The UK's early establishment of a broad-based research evaluation system provides a valuable case for analysing such long-term institutionalisation processes as well as their (un)intended consequences.

A range of studies has focused on the RAE/REF from its establishment (Barker, 2007; Sousa and Brennan, 2014) to its effects on higher education institutions (HEIs) and individual behaviour (Harley, 2002; Lucas, 2006; McNay, 1997, 2003; Talib, 2001). Other research has examined the system from disciplinary perspectives (Fisher and Marsh, 2003; Hamann, 2016; Harley and Lee, 1997), criticised its 'competitive, adversarial and punitive spirit' (Elton, 2000) and its enormous costs, discouragement of interdisciplinarity, and reduced collegiality and staff morale (Sayer, 2014)—or questioned the newer 'impact agenda' seeking to determine and enhance the relevance of research (Watermeyer, 2014). In the field of educational research, an important body of literature has focused on aspects of research assessment, such as the evolution of 'quality profiles' (Bassegy, 2002; British Educational Research Association, 1997; Gilroy and McNamara, 2009; Kerr, et. al, 1998; Lawn and Furlong, 2007), the distribution of funding through departments (Oancea, 2004), and the effects of the RAE/REF at organisational and individual levels (Furlong, 2013a; Oancea, 2010, 2014). An investigation of 'research cultures' in English and Scottish universities highlighted the highly pressured work environment and negative consequences of accountability measures (Holligan et al., 2011). Other studies discuss influences on research traditions,

resulting in the progressive homogenisation of specific fields as well as short-termism, which can threaten long-term innovation (see Knights and Richards, 2003; Lucas, 2006; Oancea, 2008). Such analyses provide salient accounts of the selective nature of the exercises, the concentration of research funding, an increasing gap between research and teaching activities, and the performativity and quality assurance cultures spreading at departmental and individual levels, with diverse positive and negative consequences, some intended, some not.

What remains less clear, however, is the impact of such changes on the behaviour of actors – whether individual academics, departments or universities – across cycles, and the possible self-fulfilling prophecies and (un)intended consequences arising from the institutionalisation of this research evaluation system for specific disciplines and research fields. Here, we address these changes on the basis of findings from a comprehensive longitudinal analysis of the field of education from 1986 to 2014.

Previous studies have tended to focus on one cycle or a comparison of specific subjects without providing a longitudinal perspective on continuity and change over a number of cycles, including results from the most recent exercise (2014). Because RAE/REF is the first and most highly institutionalised research evaluation system worldwide, the analysis presented in this article provides an invaluable longitudinal view of an influential system. Such an approach extending across decades is imperative to understand the shifts in its form as an evolving research policymaking mechanism and its (un)intended effects on submission behaviour and feedback loops via the actions of involved actors. Despite diverse perspectives on the impact of research assessment on the shape, form, and direction of educational research in the UK, there exist few systematic diachronic attempts at empirically analysing the repeated rounds of available data in the field of education.

The aims of this article are twofold. Firstly, it seeks to overcome gaps by providing a comprehensive analysis of the institutionalisation of the RAE/REF as a policy mechanism from 1986 to the most recent exercise in 2014. Seeking to counter the critique of the literature as largely ‘atheoretical’ (Lucas, 2006) or as ‘stakeholder literature’ (Gläser, 2007), our conceptual framework builds upon research literatures that focus on the dynamics of formalisation, standardisation, and transparency to explain the evolution of RAE/REF over time. Secondly, we frame RAE/REF as a measurement instrument that reflects more general trends toward ever more comparative evaluation, such as ratings and rankings, benchmarking and best practices, in an ‘age of measurement’ (Biesta, 2010) within the ‘audit society’ that aims to turn science into an auditable object (Power, 1997). We understand research evaluation systems as part of the audit society, since they share the same rationale of financial accounting that spreads and entrenches values and techniques aiming to foster enhanced performance and achieve efficiency in science, turning qualities into numeric forms such as ratings and rankings. Similarities in shared values and norms can be found in other mechanisms, such as rankings (Espeland and Sauder, 2016), international assessments (Gorur, 2014; Zapp, in press), and research evaluation that transform qualities into quantities (Espeland and Stevens, 1998) in order to measure, to compare and to inform decision making. These mechanisms result from the changing relationships between state and universities, including new forms of governance creating a ‘new social contract for science’ (Demeritt, 2000). The myriad consequences of such mechanisms are often intentional, yet many unanticipated or unintentional results also follow.

Thus, we analyse the (un)intended consequences of the British research evaluation system in educational research. We conceptualise such consequences as mechanisms of ‘reactivity,’ following Espeland and Sauder (2007, 2016). To do so, we rely on a secondary analysis of publicly available quantitative data on the number of submitted staff, outputs and research funding as well as qualitative data from existing documents and documentation, including the RAE/REF Education Panel Reports, and Higher Education Funding Councils (HEFC) reports.

We identify and discuss key trends: a *decrease* in the number of academic staff whose research was submitted for peer review assessment (hereafter: ‘submitted staff’) and submitted outputs; an *increase* in the proportion of research articles as the preferred format of submitted research output as unintended consequences; and the progressive *concentration* of research funding in a small number of education departments as an intended consequence. In routinised but evolving assessments, organisations and individuals have established strategies to achieve the best possible results. We argue that this can be understood as a form of ‘reverse engineering,’ with actors strategically shifting their behaviour according to the changing rules of the game. The submission of research articles as the preferred output and the increasing presence of applied research and quantitative approaches can be viewed as self-fulfilling prophecies that reflect major shifts in UK educational research and contemporary broader dynamics in science systems globally. Finally, the persistent or even growing concentration of funding in a small number of leading departments appears to be a largely intended consequence of the RAE/REF, contributing to the ubiquitous ‘Matthew effect’ in science (Merton, 1968), endemic in scientific evaluations. We begin with a discussion of how research evaluation systems have transformed higher education and science, and then delve into the British case, presenting findings on the institutionalisation of RAE/REF and its myriad consequences.

Research evaluation systems transforming higher education and science

Research evaluation systems (Whitley and Gläser, 2007) or performance-based research funding systems (Hicks, 2012; Roberts, 2006) are relatively recent developments, yet they are quickly transforming research around the world. Whitley (2007) defines such systems as ‘organized sets of procedures for assessing the merits of research undertaken in publicly funded organisations that are implemented on a regular basis, usually by state or state-delegated agencies’ (6). Aiming to make scientific production more accountable and performance-oriented, these systems have the potential to sustainably change both organisational structures and cultures and individual scientific activities.

Research evaluations, varying in frequency, are often ‘highly intricate, dynamic and embedded in national research systems’ (Hicks, 2012: 251); they differ in *formalisation* and *standardisation* of procedures and in their *transparency* (Whitley, 2007). Their frequency ranges from one to six or seven years and the processes vary from informal (*ad hoc* evaluations carried out at university or departmental levels) to highly formal, involving central agencies with systematic rules and procedures. *Formalisation* refers to the increasingly formal character of evaluation systems, particularly where such systems become obligatory. *Standardisation* refers to the rise of common evaluation methods or practices. Within systems, standardisation of procedures and practices varies significantly between fields and review panels (Langfeldt, 2001; Whitley, 2007). In some cases, evaluation tends to rely strongly on academic peer-review (for example in the UK, Italy and Portugal). Other systems rely more heavily on metrics, including bibliometric approaches, such as citation analysis or the use of journal impact factors. In the majority of Eastern European countries, Flanders (Belgium), and Nordic countries (Denmark, Finland, Norway, and Sweden), the use of bibliometric indicators to assess the quality of research production has increasingly become the norm.

Another key characteristic of research evaluation systems is the level of public *transparency*. Whereas in less transparent systems, assessments are done informally and results are not publicly available, systems with highly transparent procedures implement formal procedures to select panel members according to their reputation and expertise and make results publicly available. Deliberations and specific peer-review judgements usually remain undisclosed. Funding allocations through these systems vary nationally as they often replace distribution of block grants with application of quantitative formulae involving outputs, research students (studentships), external

funding (Roberts, 2006), and more recently, the attempted assessment of research ‘impact’ (Watermeyer, 2014).

To gauge the degree of their effects on universities, researchers and the research system generally, Whitley (2007) distinguishes between *weak* and *strong* research evaluation systems. Weak systems are characterised by informality, little standardisation of procedures and practices, and limited publishing of results or official ratings. These have far less impact on funding allocations and changes tend to be incremental rather than radical. Strong systems are characterised by highly formalised sets of rules and procedures and are organised around existing disciplines and scientific boundaries. The results are ranked on standard scales and publicly available and the assessment has direct impact on funding allocation. These systems’ recent elaboration and incremental changes across cycles demand empirical evidence to uncover (un)intended consequences on research policies, universities and individual researchers.

When research evaluation systems strengthen the links between government policy and university research and provide public accountability for government funds, they strongly incentivise change in individual and organisational performance (Geuna and Martin, 2003). They may increase efficiency through competition or encourage explicit research strategies and wider dissemination and bolster investment in publicising research nationally and internationally to enhance scientific reputation (Whitley, 2007). The concentration of resources may enable certain departments (universities) to compete worldwide. On the other hand, these systems regularly impose high costs for states and indirect but substantial costs for universities (Hicks, 2012). Unintentionally, they may also reduce university autonomy, decrease innovation, reduce scientific diversity or disfavour research that addresses societal needs, all depending on what ‘counts’ and what is being counted in particular assessments.

The challenge is to demonstrate how research communities react, adjust and adapt to such systems. Over time, do they become better at ‘playing the game’? Those in more favourable positions, with access to more resources and established reputations within academic hierarchies, will likely benefit most from such systems. We expect that scientific elites quickly learn an evaluation system’s rules and norms and then strategically and tactically manoeuvre to maximise their advantage(s) as a form of ‘reactivity’ (Espeland and Sauder 2007). We analyse the UK’s RAE/REF system over three decades.

Institutionalising research evaluation in the United Kingdom (1986–2014): Formalisation, standardisation, transparency

Today’s REF in the UK derives from the first and most highly institutionalised research evaluation systems worldwide. How was this system established and how has it evolved? We chart its institutionalisation (1986–2014), analysing its rationale, as well as its evolving formalisation, standardisation and transparency (see Table 1).

Rationale

The fixation on standards in education from the late 1970s (Gay, 1981) and the increasing concern for efficacy in the public sector led, in the early 1980s, to a new generation of policies in the UK and elsewhere (Bleiklie et al., 2011), that together resulted in a ‘culture of accountability’ (Biesta, 2004) and, for research, in ‘performative accountability’ (Oancea, 2008). From the beginning, the main aim was to allocate research funding more systematically and selectively. The same rationales of effectiveness and concentration have been central in all rounds. The promotion and achievement of ‘quality’ and ‘excellence’ in research has been the cornerstone of subsequent research evaluations in the UK. Yet the definition of what counts as research has evolved over time:

Table 1. The evolution of the research evaluation system in the UK (1986–2014).

	RAE 1986	RAE 1989	RAE 1992	RAE 1996	RAE 2001	RAE 2008	REF 2014
Funding agencies	Universities Grant Committee 37 CC	Universities Funding Council 152 CC	HEFCE, SHEFC, HEFCW and DENI 72 UoA	HEFCE, SHEFC, HEFCW and DENI 69 UoA	HEFCE, SHEFC, HEFCW and DENI 69 UoA	HEFCE, SHEFC, HEFCW and DEL 67 UoA	HEFCE, SFC, HEFCW and DEL 36 UoA
Cost centres (CC) or units of assessment (UoA)							
Content of submissions	<ul style="list-style-type: none"> Research income and expenditure; University's management of resources for supporting research activities and individuals; Profile for each research area. 	<ul style="list-style-type: none"> Staff return; Students and studentships; Research grants and contracts; Publications and other public output; Statement of research plans and general observations (external recognitions, substantial consultancies, etc.) 	<ul style="list-style-type: none"> Overall staff summary (census date, 31 March 1992); Active research staff return; Publications and other output; Students and studentships; External research income. 	<ul style="list-style-type: none"> Overall staff summary (census date, 31 March 1996); Research active staff; Publications and other public output; Research students and other output; External research income; Research environment and plans; General observations and additional information. 	<ul style="list-style-type: none"> Overall staff summary (census date, 31 March 2001); Research active individuals' details; Research output; Research students and studentships; External research income; Textual descriptions; General observations and additional information. 	<ul style="list-style-type: none"> Staff details (census date, 31 October 2007); Research outputs; Research students and studentships; Research income; Research environment and esteem, individual staff circumstances, and Category C staff circumstances. 	<ul style="list-style-type: none"> Staff details (census date, 31 October 2013); Publications and other forms of outputs (1 January 2008 to 31 December 2013); Unit's approach and case studies; Research doctoral degrees awarded and research income; Research environment.
Methodology: type of peer review	Subject sub-committees	70 peer reviewers	62 panels and subpanels	60 panels and subpanels	60 panels and subpanels	15 main panels and 67 subpanels	Peer review and metrics by 4 main panels and 36 subpanels
Criteria	Decided by panels	5-point rating scale	5-point rating scale	7-point rating scale (1,2,3b,3a,4,5,5 ^a)	7-point rating scale (1,2,3b,3a,4,5,5 ^a)	5-point rating scale (unclassified, 1, 2, 3, 4, 5)	Peer review and metrics by 4 main panels and 36 subpanels Outputs – 65% Impact – 20% Environment – 15% 5-point rating scale (unclassified, 1, 2, 3, 4, 5)
Grading system	Below average, average, above average or outstanding	5-point rating scale	5-point rating scale	7-point rating scale (1,2,3b,3a,4,5,5 ^a)	7-point rating scale (1,2,3b,3a,4,5,5 ^a)	5-point rating scale (unclassified, 1, 2, 3, 4, 5)	5-point rating scale (unclassified, 1, 2, 3, 4, 5)

HEFCE: Higher Education Funding Council for England; SHEFC: Scottish Higher Education Funding Council and SFC: Scottish Funding Council; HEFCW: Higher Education Funding Council for Wales; DENI: Department of Education for Northern Ireland and DEL: Department for Employment and Learning.
Source: Authors' summary.

“Research” for the purpose of the Council’s review is to be understood as original investigation undertaken in order to gain knowledge and understanding. It includes scholarship; the invention and generation of ideas, images, performances and artefacts’ (RAE, 1992).¹

‘For the purposes of the REF, research is defined as a process of investigation leading to new insights, effectively shared. It includes work of direct relevance to the needs of commerce, industry, and to the public and voluntary sectors’ (REF, 2011).

Having innovation as the main ambition, the definition of research has shifted from the pure conquest of knowledge and novelty to a more utilitarian perspective of research linked to evidence-based policy and practice, and the pursuit of ‘what works’ in several fields of expertise (Biesta, 2007). RAE 2001 introduced the idea that research must serve society facing economic and social challenges, with the notion of ‘impact’ emerging in the RAE/REF by the late 1990s and turned into an explicit component of the evaluation of the quality of research in the latest assessment (REF 2014) — to be valued even more highly, 25%, in the next cycle, REF 2021 (REF 2016; see also Martin, 2011; Watermeyer, 2014).

Formalisation and standardisation

The system’s evolution has been marked by increasing formalisation and standardisation, with the formalisation manifested in the evaluation’s size and scope and standardisation reflected in procedures and practices.

The third assessment exercise (1992) was the first to cover the whole higher education sector throughout the UK. The Further and Higher Education Act conferred university status onto 35 polytechnics, making the evaluation much more comprehensive. As a result, major changes were made in the funding and administration of the UK higher education system, including replacement of the University Funding Council by the HEFCs for England, Scotland and Wales and the Department of Education in Northern Ireland. New universities became eligible for research funding from the 1992 RAE onwards.

Regarding the information submitted, elements about staff, publications and research environment (including information about students, scholarships, external funding and strategic plans) were relatively stable through the cycles until the last exercise (REF 2014), when additional case studies about the impact of research were requested. Despite the evaluation structure remaining more or less the same, important changes from cycle to cycle have clearly affected the number of staff and outputs submitted for assessment (see Table 1).

An issue raised about the eligibility of ‘new universities’ in the RAE 1992 was which staff members should have their research selected for assessment. Previous exercises assumed *all* staff members conducted research, but in the ‘new universities’ this was not a valid assumption. In response, the category of ‘research active’ staff was introduced to distinguish between staff involved in research and teaching versus staff solely involved in teaching. Now universities could *select* which staff members to put forward for evaluation.

A further significant change after the RAE 1996 resulted from the ‘snapshot’ approach used since RAE 1989: the total number of submitted staff is based on real numbers at the census date, regardless of how long the researcher has been affiliated with the organisation in question. From the third (1992) to the fifth (2001) RAE exercises, when a researcher switched organisations during the assessment period, the credit would go to the institution to which *s/* he was affiliated at the census date. To address such attribution issues, a new staff category – Staff A* – was introduced (RAE, 2001a) to indicate those who had transferred in the preceding twelve months.

The first two RAEs requested full publication lists and the third exercise only two publications per staff member plus quantitative information about all publications, but since RAE 1996, judgement has been based on a maximum of four publications per staff member. This decision resulted from the critiques of concerned parties claiming that a broader and more in-depth review, especially of research outputs, should judge not simply *quantity* in publication lists, but rather research *quality*.

Regarding the units of assessment, the methodology and the grading system, important shifts reflect progressive *formalisation* and *standardisation*. In 1986, only three major domains of research were assessed: non-clinical scientific research, clinical scientific research, and research in the humanities and social sciences, comprising a total of 37 ‘cost centres’. In the second cycle, *all* fields were assessed in 152 different costs centres and reduced to 72 ‘units of assessment’, divided for review by 62 panels and subpanels. Thereafter, small alterations to the framework were completed ‘to reflect concerns in certain subjects, to capture properly the current layout of the research “map” and to reflect emerging disciplines’ (RAE, 1994). The fourth (RAE 1996) and the fifth (RAE 2001) evaluations had the same structure and followed the same procedures, whereas the sixth (RAE 2008) prepared the ground for considerable standardisation of the relationship between units of assessment, methodology and criteria of panels and subpanels in the seventh (REF 2014).

In all cycles, assessment has been based on peer review and conducted by panels applying their collective knowledge and experiences in their academic field. In order to ensure greater consistency across subject areas, RAE 2008 had 15 main panels with 67 subpanels, one for each unit of assessment (RAE, 2009a). Bearing in mind the necessity to change procedures to assess the quality of research cultures instead of individuals, RAE established three main domains to be assessed: research outputs, research environment and esteem indicators (RAE, 2005). Also, sub-panels had autonomy to weight the three indicators differently, depending on members’ preferences.

After RAE 2008, the RAE was replaced by the REF. Three major changes were connected with the procedures and practices of assessment, expressing a high degree of standardisation. For the first time, across all units of assessment, the weights to assess the quality profile were standardised: outputs (65%), impact (20%), and environment (15%). A metrics system was used in several but not all units of assessment. And the number of main panels and units of assessment was substantially reduced to enable greater consistency in assessment, to reduce the number of boundaries between units of assessment (reducing the need for organisations to make tactical decisions about submitted work), and to narrow disparities in sub-panel workloads (REF, 2015).

Thus far, the rating scale used has changed four times. Starting with a simple scale from ‘below average’ to ‘outstanding’, REF 2014 instead used a 5-point rating scale (unclassified, 1, 2, 3 and 4) to identify research that meets national and international standards of excellence in terms of originality, significance and rigour. The most significant changes relate to the importance ascribed to ‘international excellence’ since RAE 1996, and especially from RAE 2008, in which the second point fully addresses international excellence. From RAE 2001 onwards, panels have to consider the arrangements to involve non-UK based experts before awarding the highest ratings (RAE, 2001a). Thus, the process has also become more internationally-oriented over time.

Transparency

Progressively, the RAE has increased *transparency* in its rules, also to address criticism about procedures and mechanisms. The first exercise was considered by many as ‘quick and dirty’ (Jones and Sizer, 1990) using a ‘light touch approach’ (Gilroy and McNamara, 2009), with numerous unintended potentially deleterious consequences (Elton, 2000). Indeed, the Universities Funding Council’s (UFC) own evaluation concluded that ‘publication data was found to be unreliable, and

where it was reliable, it said nothing about the quality of the output' (Universities Funding Council, 1989, para. 23). Criticism focused on the fact that criteria for assessing research quality were not made clear to departments; that different assessment standards were being used for different subjects; and that the information sought biased judgements in favour of larger departments. Addressing such criticisms, the UFC gathered the voices of interested organisations to improve evaluation methodologies and procedures.

Nevertheless, in both RAE 1992 and RAE 1996, external and internal critiques emphasised the preference for basic research over applied research; the separation of research from teaching activities; and panel favouritism (benefitting particular organisations). After broad public consultation in 1997, for RAE 2001, all submitted information was made publicly available. Despite the efforts to make it a 'fairer' exercise, the peer-review system and the weight that some panel members placed on journal reputation and rankings and the number of citations, especially of articles, continued to raise criticism.

Despite the unavoidably subjective nature of peer review and given panel autonomy to establish quality criteria, panel reviewers have been repeatedly criticised for preferring certain fashionable topics (Morgan, 2004). In the 2003 *Review of Research Assessment*, performance indicators within a metrics system were considered, although reviewers were 'convinced that the only system which will enjoy both the confidence and the consent of the academic community is one based ultimately upon expert review' (Roberts, 2003: 7). In 2006, the government proposed the replacement of peer review by a metrics-based system for assessing and subsequently funding research in order to reduce the administrative burden associated with the exercise and meet the criticism regarding subjective bias inherent to peer review processes. Yet the metric system was only used in some fields in the natural sciences, not in all. Consultations and a pilot study were conducted to establish key parameters for the future REF. This process resulted in several changes made to satisfy criteria of equity, equality and transparency. However, a media analysis covering 341 articles (2011–2013) by Murphy and Sage (2014) concluded that discussions about REF were mainly negative, with debates dominated by the notions of 'impact', 'funding', and 'marketisation'.

In terms of the concepts proposed by Whitley (2007), therefore, over the years the UK's research evaluation system has gained in strength, is characterised by highly formalised sets of rules and procedures, is organised around existing disciplines and scientific boundaries, with results transparently ranked on standard scales that are publicly available, and where assessment has direct impact on funding allocations. It took about three decades for research assessment in the UK to develop, evolving into a formalised, standardised and highly transparent system, becoming a more or less accepted part of the research landscape. This manifests the qualities of the evaluation system as such. Yet what consequences has this trajectory towards a strong evaluation system had on the shape and form of particular disciplines, specifically educational research?

Research evaluation in the UK and (un)intended consequences: Between strength and 'reactivity'

Several authors provide insights on the effects of research evaluation systems on research policies, universities, and individual researchers (see Geuna and Martin, 2003; Whitley, 2007; Hicks, 2012). On RAE/REF a growing set of studies have pointed at several organisational, departmental and individual responses. At the organisational level, departments started to receive explicit direction to create strategies and plans for the assessment exercise (see McNay, 1997; Lucas, 2006). Translated to departmental and individual levels, research shows positive responses (e.g. the creation or strengthening of research cultures) and negatives ones, such as the increase of monitoring, pressure, and a sense of alienation and anomie (McNay, 1997; Harley, 2002; Lucas, 2006).

For educational research, in particular, an important set of studies (British Educational Research Association, 1997; Kerr, et. al, 1998; Oancea, 2004, 2010, 2014) show considerable effects of RAE/REF manifest in one or two exercises (discussed below). While these studies provide significant insights, they do not, as such, provide a comprehensive diachronic analysis of the relationship between increasing institutionalisation and (un)intended consequences for educational research.

Utilising Whitley's (2007) conceptual tools, we can confirm the strong character of the UK's research evaluation system, especially as greater levels of formalisation and standardisation were achieved. But is it the case that the stronger the research evaluation system is, the more forceful its consequences are? How have organisational and individual actors reacted to the evolving research evaluation system? We assume that the system triggered forms of reaction that have deepened or altered behaviour throughout its institutionalisation process, and we explore this assumption by analysing submission behaviour of UK Departments of Education across RAE/REF rounds.

To understand the extent of such consequences, examining feedback mechanisms triggered in the evolving institutionalisation may be helpful. According to Espeland and Sauder (2007:11), mechanisms of 'reactivity' can be understood as 'patterns that shape how people make sense of things'. As part of the 'audit society' (Power, 1997) and its increasingly routinised forms of evaluation, observation and measurement have not only become commonplace, but also call forth reactions among actors to succeed, given the rules of the game—or indeed to try to change those rules. To understand such evolving structure and action, and to uncover potential (un)intended consequences in the case of educational research, requires a longitudinal analysis. We focus on (1) 'reverse engineering' understood 'as the process of working backwards from an object in order to understand how something works' (Espeland, 2016: 281; Espeland and Sauder, 2007, 2016) and on (2) 'self-fulfilling prophecies,' as 'processes by which reactions to social measures confirm the expectations or predictions that are embedded in measures or which increase the validity of the measure by encouraging behaviour that conforms to it' (Espeland and Sauder, 2007:11).

In the next section we present the methods and data used to analyse the specific effects of RAE/REF on the field of educational research.

Methods and data to examine the effects of evaluation on educational research

In order to determine the degree of institutionalisation of the evaluation system and identify its effects, a systematic and comprehensive analysis of available data across successive cycles is crucial. To understand the behaviour of individuals and organisations, particularly with regard to the selection of submitted publications, we used data from the education unit of assessment (UoA) submissions to three research assessment cycles, RAE 2001 ($n=82$), RAE 2008 ($n=81$) and REF 2014 ($n=76$). Together, they amount to a total of 239 submissions, which, for analytical purposes, were divided into submissions from pre-1992 and post-1992 universities (see below). Examining three different dimensions – staff, research outputs (publications), and funding – we identify continuity and change in the domain of educational research compared to trends across the whole range of UoAs. For staff,² we examined total staff numbers and full-time equivalents (FTE). Within the education submissions, we also compared the total number of staff submitted for assessment of the universities to other UoAs (and in particular examined comparatively the broader group of Social Sciences and Humanities).

For external research income, we focused on differing sources of funding: UK Research Councils, UK central government bodies, UK-based charities, UK industry, and European funding. In addition, we complemented our dataset with the crucial 'quality-related' research funding for each university Department of Education from 2002 until 2015 by analysing Research Grants reports from the three HEFCs (England, Scotland and Wales) and from the Department of

Employment and Learning in Northern Ireland. It is this quality-related funding that is distributed on the basis of the outcomes of the RAE/REF. Thus, this source reflects the direct impact of research assessment on the research funding allocated to universities. (Whether, and if so, how funding is distributed to departments, is at the discretion of universities, and patterns differ widely, from full allocation to departments' budgets to no allocation to departments at all.)

For output formats, we distinguish five types of publications: books, journal articles, book chapters, conference proceedings and research reports. In order to support our claims based on the analysis of quantitative data, we also conducted a review of existing qualitative studies on the assessment and evaluation of educational research and analysed official documents such as the Education Panel Reports and HEFC reports. As previous studies indicate, increasing formalisation, standardisation and transparency entail increasing awareness and reactivity of organisations and individuals, leading to changing patterns of selectivity and submission behaviour within the departments of education. In line with criticism of the system and following the results of studies that account for the impact of RAE/REF on the choice of research topic and research publication (Kerr, et. al, 1998; Talib, 2001), we expected to find peer-reviewed articles as the preferred output for scientific evaluation, to the detriment of other publication formats. Finally, taking into account the policy of funding bodies in the UK aiming to improve research output by focusing resources on the best research performers (Department for Education and Skills, 2003), we also expected to find a growing concentration of funding in a small set of departments, contributing to the so-called 'Matthew effect' in science.

To evaluate these expectations, we used descriptive statistical analysis on the staff, funding, and outputs overall, with particular attention to the top 10 departments of education (in terms of staff size, output volume and quality profile). For all departments, we analyse variation across cycles. Moreover, to address the concentration hypothesis, we calculate a Gini index, considered to be one of the best single measures of in/equality in the distribution of resources.³

Educational research in the UK's research evaluation system

Educational research resized: Selectivity as a form of 'reverse engineering'

Educational research is currently the second largest field in the social sciences in the UK in terms of staff (Furlong, 2013b). In the last two assessment cycles, Business and Management Studies had the highest number of staff submitted for assessment, overtaking Education in the last two rounds. Education is in fact the only UoA that has continued to decrease in submitted number of staff across evaluations. As Figure 1 shows, the total number of staff submitted for assessment (Category A⁴) has decreased over the last 25 years, with a peak in 1996, and a drop of almost 50% from 1996 to 2014. Two interrelated aspects may help to explain this decline. The first is linked to historical and institutional factors of the UK's higher education system, while the second reflects the effects of the research evaluation system.

When looking at the distribution of submitted staff by type of organisation, the dominance of older universities in relation to post-1992 universities becomes visible. Historically, older universities have stronger, research-intensive profiles, while the former polytechnics focused on teaching and applied knowledge. According to Oancea (2010), the evaluations are perceived differently within these organisations. While post-1992 universities see the exercises as a means of improving their standing and of proving that they are 'REF-able', pre-1992 organisations rely heavily on the quality-related funding to develop human and material resources. Harley (2002) claims that while the RAE helped post-1992 institutions to establish a research culture, in the pre-1992 universities this pushed researchers to publish in certain refereed journals considered of high(er) quality. A

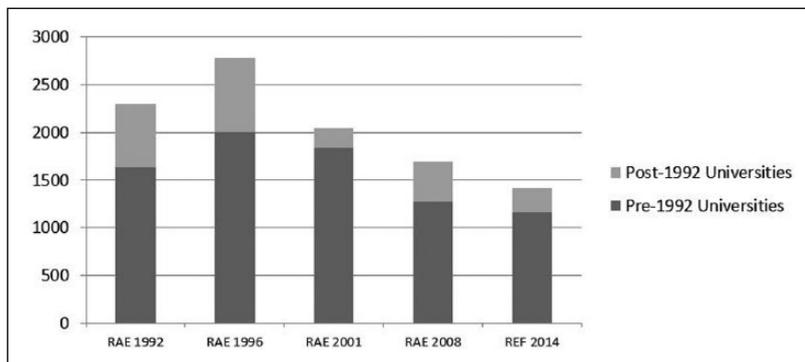


Figure 1. Staff submitted for evaluation by pre-1992 and post-1992 Universities in Education UoA, RAE 1992-REF 2014.

Source: Authors' calculations based on EDRESGOV project database (RAE 1992 to REF 2014).

review of RAE results between the third and fourth exercises emphasised reviewers' preferences for more theoretical or basic/fundamental research (Bassegy and Constable, 1997; Gilroy and McNamara, 2009; Kerr et al., 1998; Oancea, 2008). This helps explain the low results of post-1992 universities in comparison with pre-1992 universities in the first two national exercises. Given the fact that departments rated less than 3 received no quality-related funding, the post-1992 universities could hardly develop stable and strong research cultures; they were structurally disadvantaged (Alldred and Miller, 2007).

Another crucial finding is that very few departments of education submit the majority of researchers to be evaluated. In fact, only ten such departments in the UK contribute nearly half of the total submitted staff in this field (45% in RAE 2001; 49% in RAE 2008; 46% in REF 2014). Among the HEIs participating in these three cycles, just one continued to grow in staff numbers submitted. The others, especially traditional research universities, experienced a considerable drop in staff submitted to the exercise between 2001 and 2014. Despite the dominance of pre-1992 universities in the evaluation system and the concentration of almost half of submitted staff in such a small number of departments, the decrease of submitted staff is marked (Figure 1).

Comparing the 1989, 1992, and 1996 cycles, the British Educational Research Association (1997) found that the highest-rated departments had become much more selective. In fact, the average number of submitted staff per education department dropped from 23 in RAE 1996 to just 12 in REF 2014.⁵ Though these departments tend to have a greater proportion of teaching-only staff (Furlong, 2013b), the data show that over time they became more selective in assessment submissions (see Table 2). Increasingly, from RAE 2008 to REF 2014, organisations opted not to participate, as 15 did not submit in the seventh exercise. In contrast, nine organisations returned to the evaluation system after a hiatus.

Some literature suggests that over time, individuals, departments and universities have become more strategic; they have become better at playing the research assessment 'game' (Lucas, 2006; McNay, 2003). This 'playing the research game' can be understood as a mechanism of 'reverse engineering', since the increasing performativity and managerial values and their effects on the departmental and individual levels seem to have led to changes in research groups and themes, publications and output strategies, and the attraction of external funding that has, over time, increased selectivity patterns seen in the education UoA submissions (see

Table 2. Total education staff and education staff submitted for evaluation, 1995–2013.

	Total staff in education (teaching and research or research contract)	Submitted staff in education category A (RAE/REF)	Proportion of staff in education submitted
1995/96	5771	2790	48%
2000/01	5695	2045	36%
2007/08	6060	1696	28%
2013/14	5830	1442	25%

Source: Authors' calculations based on EDRESGOV project database (RAE 1996, RAE 2001, RAE 2008, REF 2014) and data from the Higher Education Statistics Agency.

Holligan et al., 2011; Oancea, 2010, 2014). The strengthening of research groups by strategically hiring new staff occurs with upcoming evaluations in mind, the progressive distinction between teaching and research activities, and elaborated internal mechanisms for peer-review and individual progress (Oancea, 2010, 2014) are all reactions to the evolving system by the departments and individuals in efforts to maximise performance.

Educational research reshaped: Output formats as self-fulfilling prophecy

In the Economic and Social Research Council review (Mills et al., 2006), education was considered among the most complex social science disciplines, which reflects the multidisciplinary character of the field in the English-speaking world (Biesta, 2011; Lawn and Furlong, 2009). Our analysis of the education panel reports (2001–2014) highlights changes over time, with specific themes becoming more ‘fashionable’ in one cycle and less in others. This is the case with adult education, lifelong learning and formal post-compulsory education in RAE 2001 (RAE, 2001b), evidencing the importance of this issue in policy agendas and the rise of lifelong learning as a global ‘paradigm shift’ in the 1990s (Zapp and Dahmen, 2017). In the next cycle, lifelong learning received less attention; instead, other themes stand out, such as service provision for children, citizenship and globalisation, and higher education research (RAE, 2009b).

Consistently, we find rising valorisation of applied research and quantitative approaches. In 2001, ‘a large proportion of research output to RAE 2001 was policy related and indicated a commitment on the part of the research community to the analysis and improvement of policy’ (RAE, 2001b). By RAE 2008, the education panel acknowledged a greater focus on applied research in such topics as linguistics or language education; class size; analyses of the value of ICT in teaching; and multimodal analysis of classroom interaction, among others. In RAE 2001, few outputs used quantitative approaches, such that ‘there is room for more approaches that use advanced quantitative methodologies’ (RAE, 2001b). During the sixth exercise (RAE 2008), more quantitative analysis and longitudinal studies were submitted, according to the education sub-panel, overcoming ‘weaknesses noted in 2001’ (RAE, 2009b). Nevertheless, ‘more longitudinal data-sets are needed, for example to provide sound evidence of long-term effects of different factors and innovations on educational outcomes’ (RAE, 2009b). This progressively stronger emphasis on quantitative approaches considerably impacted the latest submissions. As the panel stated:

Compared to 2008, significantly more outputs were submitted based on structured research designs and quantitative data. The growing volume of outputs deriving from large-scale datasets and longitudinal cohort studies was particularly impressive, and a high proportion were judged to be internationally excellent or world-leading (REF, 2015).

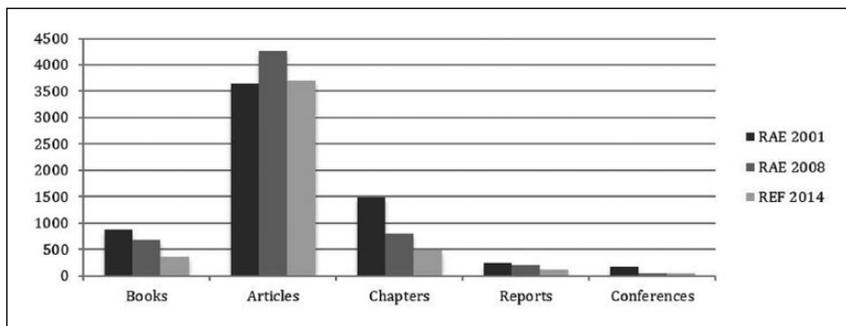


Figure 2. Total submitted publications by type, 2001-2014.

Source: Authors' calculations based on EDRESGOV project database (RAE 2001, RAE 2008, REF 2014).

This noted move deserves closer analysis given the field's internal dynamics over the last decades. Previous studies found a correlation between the subjects and the rating of departments, affirming that issues relating to education policy and governance (Bassey and Constable, 1997; Kerr et al., 1998) and topics more related to basic research were present in departments with higher ratings. Oancea (2004) found similar results in RAE 2001, although the presence of longitudinal research, methodological issues, economics and educational psychology started to figure more prominently in the research portfolios of the higher-rated departments.

In light of criticisms that the system was privileging basic research over applied research, it seems that progressively more applied research has been submitted, especially over the last two rounds, possibly due to increases in funding, for example, from the Teaching and Learning Research Programme (TLRP), whose origins are linked to responses to the perceived educational research crisis in the 1990s (Hillage et al., 1998; Tooley and Darby, 1998).⁶ TLRP was conceived to reclaim a space for research to inform practice and policy and to improve the quality of educational research (see also Biesta, 2016). In 2014, early childhood education, higher education and teaching and learning were once again highlighted because of TLRP, along with considerable growth of evaluation boosted by the Education Endowment Foundation (REF, 2015).

Examining research outputs specifically, we observe that there was a decrease through the years in all forms of publications submitted, except in RAE 2008 in journal articles (see Figure 2). Despite the decrease registered across output forms, articles gained considerably. If in 2001, 57% of total outputs was in article form, by 2008 this was 71% and in 2014 77% – nearly four-fifths. Evidencing the importance of research outputs, these were weighted to 70% by education panels (2008) and 65% by all panels (2014).

The expanding presence of quantitative studies can be understood as part of the broader dynamics of public research funding instruments expressed in the last three rounds of the exercise – after the government's strategic investment to improve the quality of educational research, such as the case of TLRP. We consider that the visible trend toward articles as the preferred output can be understood as a self-fulfilling prophecy throughout the evaluation system's institutionalisation. In educational research, all forms other than articles – books, chapters, reports, and conference proceedings – declined, especially the first two. The article has long been the gold standard of scientific production in many fields, and production been expanding exponentially worldwide over the past few decades (see Powell, Baker and Fernandez, 2017). In many fields, advancing bibliometrics and evaluations that emphasise impact factors and citations have further strengthened the importance of journal articles for career progression, to gain individual and organisational

reputation, and to inform funding agencies ex-ante about the proven capacity to conduct research. Within such global dynamics, the RAE/REF specifically augmented the emphasis on the production of research articles. Clearly, this trend extends across many fields and contexts. Through a longitudinal analysis of UK science covering almost 20 years, Moed (2008) found several distinct bibliometric patterns. In RAE 1992, UK scientists substantially increased their article production, while in RAE 1996 authors increased the number of publications in journals with a high citation impact, and, from 1997 to 2000, departments stimulated their staff members to collaborate more intensively. In Europe, similar growth patterns in the publication of articles can be found in Eastern European countries (Pajić, 2015), Spain (Jimenez-Contreras et al., 2003), Norway (Aagaard, Bloch and Schneider, 2015), Belgium (Flanders) (Engels et al., 2012), and Portugal (Viseu, 2016) since the implementation and strengthening of research evaluation systems.

Not only productivity in general matters, but also what kinds of outputs are preferred. In the study of the impacts of RAE 2008, Oancea (2010, 2014) found that individuals feel pressured to produce four mainstream outputs in particular formats, such as journal articles. Other education-related studies have discussed the evaluation system as an external tool used internally by university managers to steer publication strategies, a clear example of reactivity. In their study of higher education researchers' perceptions and experiences, Leathwood and Reed (2012) found encouragement and internal and external pressures or even 'surveillance and/or threats' placed on academics to perform for RAE/REF, significantly influencing career progression. This reported pressure seems more pervasive in older than in newer universities.

The RAE 2008 education panel clearly stated: 'A move was noted from book to journal publishing; the trend may be perceived as encouraged by RAE requirements' (RAE, 2009b). Even without regulative guidelines for article submission, the normative pressure emphasises reliance on the already-proven quality of peer-reviewed pieces of scholarship. Recognising the widespread assumption that articles are evaluated as the most valuable form upon which reputation and funding are distributed, academics reinforce the tendency as they choose to write more articles and, in turn, submit them more frequently (and to particular journals), leading to a self-fulfilling prophecy. Yet not only reputation, credibility and prestige align behaviour to the research evaluation system, as the provided funding includes concrete incentives.

Funding for educational research: Progressive concentration explicitly intended

Educational research in the UK is funded from two main sources. External funding comes from a range of different bodies, such as the research councils, UK government, industry, charities, and the EU, and is awarded on the basis of competitive tendering. Core funding for research is given to universities on an annual basis through the HEFCs in England, Scotland and Wales and through the Department for Employment and Learning in Northern Ireland. This funding is known as 'quality-related research funding' (QR) and is annually distributed according to the evaluated achievement of each department in the RAE/REF, although how achievement is 'translated' into funding has evolved over time, even between assessment cycles.

Figure 3 depicts the different sources of funding to education departments.⁷ From 2002 to 2008, external sources provided 60% of funding, while in the next period this rose to two-thirds. Only 40% (2002–2008⁸) and 35% (2009–2014) of funding to departments of education is distributed through the research evaluation system. Public bodies, namely the research councils and the UK government, provide most of the funding (62%, RAE 2001; 79%, RAE 2008, and 69%, REF 2014), with these fluctuations relating to two particularly important events for educational research. The increase in the period between rounds RAE 2001 to RAE 2008 reflects the £43m channelled through the TLRP from 2001 to 2011. Consequently, the end of this programme but

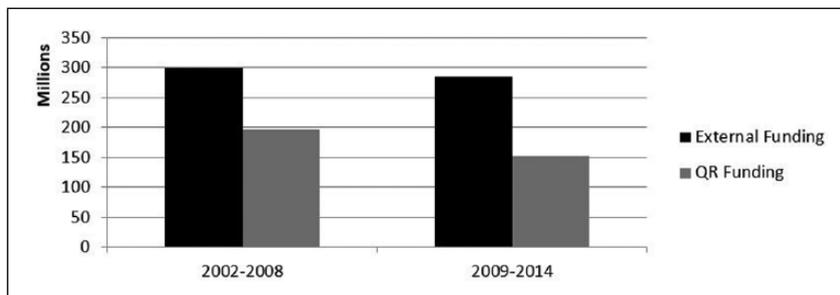


Figure 3. External and quality-related funding of education departments, UK (2002–2014).

Source: Authors' calculations based on EDRESGOV project database (RAE 2001, RAE 2008, REF 2014) and HEFCE: Higher Education Funding Council for England; SFC: Scottish Funding Council; HEFCW: Higher Education Funding Council for Wales; DEL: Department for Employment and Learning.

also the cancellation of 13 projects in 2010 (around £7m) and another 14 projects before a contractor for the full research design (Whitty et al., 2012) led to a funding decrease (2008–2014). Fluctuations in private/charitable funding (28%, RAE 2001; 16%, RAE 2008; and 23%, REF 2014) and in EU funding (10%, RAE 2001; 5%, RAE 2008; and 8%, REF 2014) also exist across cycles. Public funding is the first source of funding, with educational research remaining highly dependent on it. While the distribution of external funding does not show a tendency of concentration between cycles, previous literature showed a correlation between the capacity to attract external funding and received ratings (in line with our results). The considerable concentration of resources – almost half – in only ten departments was maintained for all three analysed cycles. The distribution of external funding among educational departments from RAE 2001 to RAE 2008 shows significant funding increases allocated to very few departments. There may be a tendency over the years for this funding to become more concentrated as external funding totals in REF 2014 rose above those in RAE 2001. Our calculations show, in fact, that the distribution of external funding overall on the basis of REF 2014 results (65.8 Gini coefficient) is more concentrated at departmental level than it was in RAE 2001 (53). Taking the top-ten ranked departments in the last three exercises shows a clear pattern of concentration of external funding (38%, RAE 2001; 46%, RAE 2008; 53%, REF 2014): this illustrates the policy-induced consequence of the top organisations being able to maximise their participation and results.

Regarding the QR funding for research, each HEFC in the UK has different rules and formulae to distribute its research grants to universities, and, as mentioned, the rules have also changed between assessment cycles. For instance, England's Funding Council decided to divide the total available funding between the subject fields of all panels in proportion to the assessed volume of research in each field. Then, the volume is measured by multiplying the number of research-active staff employed by the proportion of research that meets a certain quality threshold (grade point average). On this basis, QR funding is allocated. For educational research UK-wide, the total amount of QR funding has decreased since the academic year 2006/07 from around £31m to around £24m in 2015/16. The largest amount of QR funding for educational research comes from HEFC for England (84%, academic year 2015/16), followed by Scotland (11%), Northern Ireland (3%), and Wales (2%).

Not only in education does the UK's research evaluation system seem to be getting increasingly selective in the allocation of public funds for research (see Table 3). With the 2003 White Paper *The Future of Higher Education*, the UK government clearly expressed the desire to further concentrate public funding: 'we also need to reap the benefits which flow from concentrating the best research in larger units' (Department for Education and Skills, 2003: 5). Following this publication, HEFCE revised the allocation of QR funding and for academic year 2003/2004 ceased to fund departments

Table 3. Distribution of external and quality-related funding, UK (2002-2014).

	RAE 2001					RAE 2008					REF 2014				
Total external funding	53.04					73.98					65.77				
QR	02/03	03/04	04/05	05/06	06/07	07/08	08/09	09/10	10/11	11/12	12/13	13/14	14/15	15/16	
	55.87	49.90	51.75	50.47	50.82	50.81	51.18	58.06	58.99	60.37	60.83	60.85	60.85	83.60	

Note: The Gini index of concentration or coefficient is used to measure distributional inequalities; here the coefficient is expressed as an average; own calculations; 0 = complete equality, 100 = complete inequality.

Source: Authors' calculations based on EDRESGOV project database (RAE 2001, RAE 2008, REF 2014) and HEFCE: Higher Education Funding Council for England; SFC: Scottish Funding Council; HEFCW: Higher Education Funding Council for Wales; DEL: Department for Employment and Learning.

rated below 4, while in the previous exercise those rated with 3a received some funding. This policy-driven concentration of funding is noticeable also in the two most recent exercises. After RAE 2008, departments rated with 4*, 3*, and 2* were funded according to the weighting: 4*/7, 3*/3, and 2*/1, while after REF 2014 universities were only funded for the percentage of 3* and 4* submitted work. The goal of policymakers, implemented by the HEFCE, to increase the selection and concentration of funding is manifest in our results: we find considerable disparities between QR funding distribution after REF 2014 (83.6 Gini coefficient; academic year 2015/2016) and the average in RAE 2001 (51.5 Gini coefficient) and RAE 2008 (59.9 Gini coefficient).

Furthermore, those educational departments that managed to perform better and those with the capacity to attract external funding are often the same, corroborating the results of Kerr et al. (1998) on the relationship between external funding and RAE grading. Oancea (2004, 2010), in the British Educational Research Association report of RAE 2001 and RAE 2008 results, found a relationship between ratings and external funding, in which departments rated 3a and under seemed in a less advantaged position to attract UK Government funding. This concentration might be explained by the capacity of better-ranked education departments to attract external funding based on their human and material resources and reputation in the UK and beyond, reinforcing the 'Matthew effect' in science.

Discussion and conclusions

Our analysis of the institutionalisation of the UK's research evaluation system from 1986 to 2014 confirms that research assessment has been firmly established within UK higher education as a strong system, increasingly formalised and standardised. Increasingly transparent, the research evaluation system has had various (un)intended consequences. We found impact on educational research with regard both to this multidisciplinary field's structural organisation and its cognitive development. While the existing literature clearly documents the widespread perceptions of RAE/REF and diverse effects on departments and individuals, studies examining submission behavior of departments of education charting the gradual strengthening of the evaluation system, especially after RAE 2008, were lacking.

The HEFC's concentration of funding where excellence is found led educational departments to become more selective in their submissions, as strategies to improve departmental research performance developed. The overall 'quality profile' of educational research improved across cycles as participants in the system gained knowledge about and experience with this evaluation system, reacting to incentives and gradual formalisation and standardisation.

We conclude that as the research evaluation system became more formalised, standardised and transparent, both departments and individual scholars found ways to maximise their 'research

power' and their positioning via the rating scale. While individuals seek to enhance their productivity, departments seek to maximise their performance, choosing among its research workforce or contracting individuals with the highest 'research power' prior to the evaluation. Indeed, the successive rounds of the RAE/REF have provided windows of opportunity to conduct 'reverse engineering' as individuals have begun to better understand the rules of the game. The reactivity triggered by on-going formalisation and standardisation – coupled with transparency – of the system enabled departments and individuals to learn how to maintain or enhance their position in an always competitive, stratified academic world.

Among the effects on educational research as a multidisciplinary field, we have found incremental shifts toward quantitative and applied research. These reflect academic and social claims for more evidence to inform practice and policy via research of relevance and demonstrable social 'impact'. The demand for enhanced quality of research particularly reflects the UK's educational research landscape in the 1990s that witnessed major public research funding via the TLRP. With concrete, longitudinal evidence, we found that the institutionalisation of the research evaluation system has affected the publication outputs submitted for review – peer-reviewed journal articles – or at least has reinforced specific behaviour in submission procedures. Since the system was not originally designed with an explicit preference for journal articles, this can be seen as a self-fulfilling prophecy, but also as being linked to larger trends in scientific productivity found in other European countries, such as Germany and France, and in other disciplines (Powell and Dusdal, 2017). It may also be understood as a largely unintended consequence of the research evaluation system's emphasis on formats most amenable to peer review.

In terms of funding, we found modest increases in the last two exercises in external funding by public and private funding streams, and fluctuations in funding from the UK government. On the other hand, we found a decrease in funding for educational research after 2008. Relating to both external and QR funding, our research confirms other studies that suggest that the UK research evaluation system contributes to the 'Matthew effect' in science, as the strongest departments not only participated continuously, but achieved top ratings and positions in rankings, thus maximising their ability to attract further external funding.

The impact of research policy is never unidirectional, but is rather shaped by the strategic and anticipatory actions of those affected by policies, especially regarding the scarce and positional goods of research funding, individual reputation and organisational status. Whether intended or not, RAE/REF has specific consequences, such as selectivity and concentration, as it shifts and reinforces certain dynamics in the contemporary higher education landscape. Research evaluation systems such as the RAE/REF do not simply shape the structural organisation and cognitive development of research fields, but induce complex dynamics as much influenced by the workings of policy as by the 'uptake' of policy by individuals and organisational actors. Policymakers in other countries should be aware of the intended and unintended consequences shown here as they model and reform their evaluation systems and attempt to measure the results of their investments in educational research.

Declaration of Conflicting Interest

The authors declare that there is no conflict of interest.

Funding

This research was funded by the University of Luxembourg's tandem programme for interdisciplinary research (Project: The New Governance of Educational Research. Comparing Trajectories, Turns and Transformations in the United Kingdom, Germany, Norway and Belgium, EDRESGOV).

Notes

1. Because the older RAE documents are Internet-based, they contain no page numbers.
2. The number of submitted staff is publicly available for RAE 1992 and RAE 1996. We conducted secondary analysis of data on staff from the Higher Education Statistics Agency.
3. A Gini index of concentration or coefficient is used to measure distributional inequalities. We apply this tool here to show how (un)equally the funding for education has been distributed; the higher the number above 0, the higher the inequality. Were one department to receive all funding (total inequality), the score would be 1.0.
4. Category A corresponds to the total full-time equivalents and is defined as academic staff with an employment contract of 0.2 full-time equivalents or greater and on the payroll of the submitting HEIs on the census date, and whose primary employment function is to undertake either 'research only' or 'teaching and research'.
5. For accounts of UK's educational research landscape overall, see Whitty et al. (2012) and Oancea and Mills (2014).
6. For a detailed analysis of the educational crisis, see Oancea (2005).
7. For the period 2002 to 2008, we gathered all the external funding in the RAE 2008 submissions that corresponds to the external research income (from 2002 to 2007) and gathered the QR funding for each academic year, which was distributed based on the results of RAE 2001. For the next period, we follow the same approach. We compare the funding awarded by external sources received from 2009 to 2014 present in education department submissions in REF 2014 with the QR funding for each academic year that was distributed based on the results of RAE 2008.
8. Data for Scotland and Northern Ireland were not available from the academic years 2002/2003 to 2005/06. The authors contacted the designated bodies and both confirmed that this data was not available.

References

- Aagaard K, Bloch C and Schneider J (2015) Impacts of performance-based research funding systems: The case of the Norwegian Publication Indicator. *Research Evaluation* 24(2): 106–117.
- Allred P and Miller T (2007) Measuring what's valued or valuing what's measured? Knowledge production and the Research Assessment Exercise. In: Gillies V. and Lucey H. (eds) *Power, Knowledge and the Academy*. Basingstoke, UK: Palgrave, pp.147–67.
- Barker K (2007) The UK Research Assessment Exercise: The evolution of a national research evaluation system. *Research Evaluation* 16(1): 3–12.
- Bassey M (2002) RAE 2001: Analysis of the education results. *Research Intelligence* 78(February): 8–14.
- Bassey M and Constable H (1997) Higher education research in education, 1992–1996: Fields of enquiry reported the HEFCs RAE. *Research Intelligence* 61(July): 6–8.
- Besley T and Peters M (2009) Neoliberalism, performance and the assessment of educational research quality: Comparing United Kingdom, Australia and New Zealand. In: Besley T (ed.) *Assessing the Quality of Educational Research in Higher Education. International perspectives*. Rotterdam, The Netherlands: Sense, pp.27–48.
- Biesta G (2004) Education, accountability and the ethical demand: Can the democratic potential of accountability be regained? *Educational Theory* 54(3): 233–250.
- Biesta G (2007) Why 'what works' won't work: Evidence-based practice and the democratic deficit in educational research. *Educational Theory* 57(1): 1–22.
- Biesta G (2010) *Good Education in an Age of Measurement: Ethics, politics, democracy*. Boulder, CO: Paradigm Publishers.
- Biesta G (2011) Disciplines and theory in the academic study of education: A comparative analysis of the Anglo-American and continental construction of the field. *Pedagogy, Culture and Society* 19(2): 175–192.
- Biesta G (2016) Improving education through research? From effectiveness, causality and technology, to purpose, complexity and culture. *Policy Futures in Education* 14(2): 194–210.
- Bleiklie I, Enders J, Lepori B and Musselin C (2011) New public management, network governance and the university as a changing professional organization. In: Christensen T and Laegried P. (eds) *The Ashgate Research Companion to New Public Management*. Farnham, UK: Ashgate, pp.161–176.

- British Educational Research Association (1997) Results of the Research Assessment Exercise. *Research Intelligence* 59(February): 4–7.
- Demeritt D (2000) The new social contract for science: Accountability, relevance and value in US and UK science and research policy. *Antipode* 32(3): 308–329.
- Department for Education and Skills (2003) *The Future of Higher Education*. London: HMSO.
- Elton L (2000) The UK Research Assessment Exercise: Unintended Consequences. *Higher Education Quarterly* 54(3): 274–283.
- Engels TCE, Ossenblok TLB and Spruyt EHJ (2012) Changing publication patterns in the social sciences and humanities, 2000–2009. *Scientometrics* 93(2): 373–390.
- Espeland W (2016) Reverse engineering and emotional attachments as mechanisms mediating the effects of quantification. *Historical Social Research* 41(2): 280–304.
- Espeland W and Sauder M (2007) Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology* 113(1): 1–40.
- Espeland W and Sauder M (2016) *Engines of Anxiety: Academic rankings, reputation, and accountability*. New York: Russell Sage Foundation.
- Espeland W and Stevens M (1998) Commensuration as a social process. *Annual Review of Sociology* 24: 313–343.
- European Commission (2010) *Assessing Europe's University-Based Research*. Brussels, Belgium: European Commission.
- Fisher M and Marsh P (2003) Social work research and the 2001 Research Assessment Exercise. *Social Work Education* 22(1): 71–80.
- Furlong J (2013a) Research assessment and the shaping of educational research in the UK. *The Australian Educational Researcher* 40(4): 513–520.
- Furlong J (2013b) *Education. An anatomy of the discipline*. Abingdon, UK: Routledge.
- Gay B (1981) Accountability in education: A review of the literature and research in the United Kingdom. *Westminster Studies in Education* 4(1): 29–43.
- Geuna A and Martin B (2003) University research evaluation and funding. *Minerva* 41(4): 277–304.
- Gilroy P and McNamara O (2009) A critical history of research assessment in the United Kingdom and its post-1992 impact on education. *Journal of Education for Teaching* 35(4): 321–335.
- Gläser J (2007) The social order of research evaluation systems. In: Whitley R and Gläser J (eds.) *The Changing Governance of the Sciences*. Dordrecht, The Netherlands: Springer, pp.245–264.
- Gorur R (2014) Towards a sociology of measurement in education policy. *European Educational Research Journal* 13(1): 58–72.
- Hamann J (2016) The visible hand of research performance assessment. *Higher Education* 72(6): 761–779.
- Harley S (2002) The impact of research selectivity on academic work and identity in UK universities. *Studies in Higher Education* 27(2): 187–203.
- Harley S and Lee FS (1997) Research selectivity, managerialism, and the academic labour process. *Human Relations* 50(11): 1427–1460.
- Hicks D (2012) Performance-based university research funding systems. *Research Policy* 41(2): 251–261.
- Hillage J, Pearson R, Anderson A and Tamkin P (1998) *Excellence in Research on Schools*. London: Department for Education and Employment.
- Holligan C, Wilson M and Humes W (2011) Research cultures in English and Scottish university education departments. *British Educational Research Journal* 37(4): 713–734.
- Jimenez-Contreras E, Anegón F and Lopez-Cozar E (2003) The evolution of research activity in Spain: The impact of the National Commission for the Evaluation of Research Activity (CNEAI). *Research Policy* 32(1): 123–142.
- Jones PK and Sizer J (1990) The Universities Council's 1989 Research Selectivity Exercise. *Beiträge zur Hochschulforschung*, 4: 309–348.
- Karlsson S (2017) Evaluation as a travelling idea: Assessing the consequences of Research Assessment Exercises. *Research Evaluation* 26(2): 55–65.

- Kerr D, Lines A, MacDonald A and Andrews L (1998) *Mapping education research in England: An analysis of the 1996 RAE*. Bristol, UK: Higher Education Funding Council for England.
- Knights D and Richards W (2003) Sex discrimination in UK academia. *Gender, Work and Organization* 10(2): 213–238.
- Langfeldt L (2001) The decision-making constraints and processes of grant peer review, and their effects on the review outcome. *Social Studies of Science* 31(6): 820–841.
- Lawn M and Furlong J (2007) The social organization of educational research in England. *European Educational Research Journal* 6(1): 55–70.
- Lawn M and Furlong J (2009) The disciplines of education in the UK. *Oxford Review of Education* 35(5): 541–552.
- Leathwood C and Read B (2012) *Final report: Assessing the impact of developments in research policy for research on higher education*. London: Society for Research on Higher Education.
- Lucas L (2006) *The Research Game in Academic Life*. Maidenhead, UK: Open University Press.
- Martin BR (2011) The Research Excellence Framework and the ‘impact agenda’: Are we creating a Frankenstein monster? *Research Evaluation* 20(3): 247–254.
- Martin BR and Whitley R (2010) The UK Research Assessment Exercise: a case of regulatory capture? In: Whitley R, Gläser J and Engwall L (eds) *Reconfiguring Knowledge Production*. Oxford: Oxford University Press, pp.51–80.
- Merton RK (1968) The Matthew effect in science. *Science* 159(3810): 56–63.
- McNay I (1997) *The impact of the 1992 RAE on institutional and individual behaviour in English higher education*. Bristol, UK: Higher Education Funding Council for England.
- McNay I (2003) Assessing the assessment: An analysis of the UK Research Assessment Exercise, 2001, and its outcomes, with special reference to research in education. *Science and Public Policy* 30(1): 47–54.
- Mills D, Jepson A, Coxon T, Easterby-Smith MPV, Hawkins P and Spencer J (2006) *Demographic Review of the UK Social Sciences*. UK: Economic and Social Research Council.
- Moed HF (2008) UK Research Assessment Exercises. *Scientometrics* 74(1): 153–161.
- Morgan KJ (2004) The Research Assessment Exercise in English universities, 2001. *Higher Education* 48(4): 461–482.
- Murphy T and Sage D (2014) Perceptions of the UK’s Research Excellence Framework 2014. *Journal of Higher Education Policy and Management* 36(6): 603–615.
- Oancea A (2004) The distribution of education research expertise: Findings from the analysis of RAE 2001 submissions (I, II). *Research Intelligence* 87(May)/88(August): 3–8 / 3–9.
- Oancea A (2005) Criticisms of educational research: Key topics and level of analysis. *British Educational Research Journal* 31(2): 157–183.
- 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 BERA/UCET Review of the Impacts of RAE 2008 on Education Research in UK Higher Education Institutions*. Macclesfield, UK: Universities’ Council for the Education of Teachers / British Educational Research Association.
- Oancea A (2014) Research assessment as governance technology in the United Kingdom. *Zeitschrift für Erziehungswissenschaft* 17(6): 83–110.
- Oancea A and Mills D (2014) *Educational Research*. Final report of the British Educational Research Association Observatory 2014. London: British Educational Research Association.
- OECD (2010) *Performance-based Funding for Public Research in Tertiary Education Institutions*. Paris: OECD Publishing.
- Pajić D (2015) Globalization of the social sciences in Eastern Europe: Genuine breakthrough or a slippery slope of the research evaluation practice? *Scientometrics* 102(3): 2131–2150.
- Power M (1997) *The audit society: Rituals of verification*. Oxford: Oxford University Press.
- Powell JJW, Baker DP and Fernandez F (eds) (2017) *The Century of Science: The global triumph of the research university*. International Perspectives on Education & Society, vol.33. Bingley, UK: Emerald Publishing.

- Powell JJW and Dusdal J (2017) The European center of science productivity: Research universities and institutes in France, Germany, and the United Kingdom. In: Powell JJW, Baker DP and Fernandez F (eds). *The Century of Science: The global triumph of the research university*. International Perspectives on Education & Society, vol.33. Bingley, UK: Emerald Publishing, pp. 55–84.
- Sayer D (2014) *Rank Hypocrisies*. London: SAGE.
- Sousa SB and Brennan J (2014) The UK Research Excellence Framework and the transformation of research production. In: Musselin C and Teixeira P (eds) *Reforming Higher Education: Public policy design and implementation*. Dordrecht, The Netherlands: Springer, pp.65–80.
- RAE (1992) Research Assessment Exercise 1992: The outcome. Circular 26/92. Available at: http://www.rae.ac.uk/1992/c26_92.html#annexa
- RAE (1994) Research Assessment Exercise: Units of assessment. RAE96 3/94. Available at: http://www.rae.ac.uk/1996/c3_94.html
- RAE (2001a) Research Assessment Exercise: The outcome. RAE 4/01. Available at: http://www.rae.ac.uk/2001/Pubs/4_01/section1.asp
- RAE (2001b) Research Assessment Exercise. Overview reports from the panels. Education. Available at: <http://www.rae.ac.uk/2001/overview/docs/UoA68.pdf>
- RAE (2005) Research Assessment Exercise: Guidance to panels. RAE 01/2005. Available at: <http://www.rae.ac.uk/pubs/2005/01/rae0105.pdf>
- RAE (2009a) Research Assessment Exercise: Manager's report. Available at: <http://www.rae.ac.uk/pubs/2009/manager/manager.pdf>
- RAE (2009b) Research Assessment Exercise. Overview report. Sub-Panel 45 education. Available at: <http://www.rae.ac.uk/pubs/2009/ov/>
- REF (2011) Assessment framework and guidance for submissions. REF 02.2011. Available at: <http://www.ref.ac.uk/media/ref/content/pub/assessmentframeworkandguidanceonsubmissions/GOS%20including%20addendum.pdf>
- REF (2015) Research Excellence Framework 2014: Overview report by Main Panel C and sub-panels 16–26. Available at: <http://www.ref.ac.uk/media/ref/content/expanel/member/Main%20Panel%20C%20overview%20report.pdf>
- REF (2016) *Building on Success and Learning from Experience: An independent review of the Research Excellence Framework*. London: Department for Business, Energy and Industrial Strategy.
- Roberts G (2003) Review of research assessment. Available at: http://www.ra-review.ac.uk/reports/roberts/roberts_summary.pdf
- Roberts P (2006) Performativity, measurement and research. In: Ozga J, Seddon T and Popkewitz T (eds) *Education Research and Policy*. London: Routledge, pp.185–199.
- Talib A (2001) The continuing behavioural modification of academics since the 1992 Research Assessment Exercise. *Higher Education Review* 33(3): 30–46.
- Tooley J and Darby D (1998) *Educational Research: A critique. A survey of published educational research*. London: Office for Standards in Education.
- University Funding Council (1989) Report on the 1989 RAE. London: University Funding Council.
- Viseu S (2016) Play the game or get played? Researchers' strategies around R&D policies. In K. Trimmer (ed.). *Political Pressures on Educational and Social Research – International Perspectives*. Abingdon, UK: Routledge, pp.55–56.
- 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(2): 359–377.
- Whitley R (2007) Changing governance of the public sciences. In: Whitley R and Gläser J (eds) *The Changing Governance of the Sciences*. Dordrecht, The Netherlands: Springer, pp.3–26.
- Whitley R and Gläser J (2007) *The Changing Governance of the Sciences*. Dordrecht, The Netherlands: Springer.
- Whitty G, Donoghue M, Christie D, Kirk G, Menter I, McNamara M, Moss G, Oancea A, Rogers C and Thomson P (2012) *Prospects for Education Research in Education Departments in Higher Education Institutions in the UK*. London: Universities' Council for the Education of Teachers / British Educational Research Association.

Zapp M (in press) Beyond large-scale achievement testing – The psychological turn in international organizations' work on educational assessment. In: A. Wiseman (ed.) *Cross-nationally Comparative, Evidence-based Educational Policymaking and Reform*. Bingley, UK: Emerald Publishing.

Zapp M and Dahmen C (2017) The diffusion of educational ideas. An event history analysis of lifelong learning. *Comparative Education Review* 61(1): 492-518.

Author biographies

Marcelo Marques is a PhD student at the Institute of Education and Society, University of Luxembourg. His research interests are in the field of comparative education with a particular focus on the Europeanization of education systems and higher education and research. His thesis focuses on the analysis of public research funding instruments, at national and international levels, and their effects on educational research in England.

Justin JW Powell is Professor of Sociology of Education, Institute of Education & Society, University of Luxembourg. Recent books include *Comparing Special Education: Origins to contemporary paradoxes* (Stanford, 2011), *Barriers to Inclusion: Special education in the United States and Germany* (Routledge, 2016), and the edited volume *The Century of Science: The global triumph of the research university* (Emerald, 2017).

Mike Zapp is a post-doctoral researcher at the Institute of Education & Society at the University of Luxembourg. His research consists of comparative institutional analyses of education systems with a particular focus on the role of international organizations and higher education. His research has appeared in such journals as *Comparative Education Review*, *International Journal of Educational Development*, *European Educational Research Journal*, and *Science and Public Policy*. His co-authored book (with M. Marques and J. Powell) *European Educational Research (Re-)Constructed: Institutional Change in Germany, the United Kingdom, Norway and the European Union* is forthcoming from Symposium Books.

Gert Biesta (www.gertbiesta.com) is Professor of Education and Director of Research at the Department of Education of Brunel University London, UK; NIVOZ Professor for Education at the University of Humanistic Studies, Utrecht, the Netherlands; and Professor II at NLA University College, Bergen, Norway. He is associate editor of the journal *Educational Theory*, and co-editor of the *British Educational Research Journal*. He writes about the theory of education and educational research. Recent books include *The Rediscovery of Teaching* (Routledge 2017) and *Letting Art Teach: Art Education after Joseph Beuys* (ArtEZ Press 2017).