Educational Management and Policy
Volume 1, Number 1, 2026, 43-68
DOI: 10.70116/31434541161
ISSN: 3143-4541
Article

Assessment bias in the 2014 UK Research Excellence Framework: Evidence from economics and business management

James Johnston [1]

[1] University of the West of Scotland, School of Business and Creative Industries, Paisley, United Kingdom

Abstract

The standing of universities is increasingly tied to assessments of their research prowess. Research assessment exercises have the potential to provide participants in Higher Education (HE) systems with objective assessments of research quality. However, if the evaluations of research quality produced by research assessment exercises are biased, this may undermine their contribution to a more efficient allocation of scarce resources, both within and between universities. If the bias takes the form of some Units of Assessment (UOA) receiving unjustifiably favourable assessments, submissions to these UOAs may be viewed more favourably by university managers when it comes to deciding where to put their scarce research funding. As well as distorting the internal allocation of resources, a research assessment process which lacks objectivity may alter a university's reputational capital, which in turn may affect its ability to attract and retain students and staff. This paper provides fresh evidence based on decomposition techniques on the merits of claims of bias in research assessment exercises using data on submissions to the Business and Management Studies (BMS) and Economics and Econometrics (E&E) UOAs at the 2014 REF.

Article History
Received 17.02.2026
Accepted 28.04.2026

Keywords
Research assessment; assessment bias; higher education policy; research evaluation

Introduction

Research assessment exercises play a central role in shaping the allocation of scarce resources within and between universities in the UK's Higher Education (HE) system. As universities have become increasingly marketised and managerialist in their orientation, the results of exercises such as the Research Excellence Framework (REF) have taken on growing importance, influencing institutional funding, reputational capital, and the recruitment and retention of both students and staff. However, the objectivity of these exercises has been questioned, with some scholars arguing that certain Units of Assessment (UOAs) benefit from more generous evaluations than others. This paper examines this question by applying Blinder-Oaxaca and related decomposition techniques to data on submissions to the Business and Management Studies (BMS) and Economics and Econometrics (E&E) UOAs at the 2014 REF, with the aim of establishing how much of the gap in mean scores between the two UOAs is attributable to differences in research quality and how much, if any, reflects bias in the assessment process.

Corresponding Author James Johnston James.Johnston@uws.ac.uk University of the West of Scotland, School of Business and Creative Industries, George St, Paisley PA1 2BE, United Kingdom

© 2026 by Symphony. This is an open access article under the CC BY 4.0 license.

Changes to the Size and Composition of the UK's HE Sector

In common with HE systems elsewhere in the world, over the last half century, the UK's system has gone through a process referred to by Trow (1973) as "massification". Until the 1980s, the UK had a small, elite HE system, funded from general taxation and "free at the point of use". Indeed, not only were there no student tuition fees, but the state provided means-tested grants to support students from lower income households with their living costs while they studied. In the 1960s, the number of universities in the UK doubled from 22 to 45 as a result of the Robbins report (1963), which brought into being the chartered universities. Thirty years later, the Further and Higher Education Act of 1992 allowed forty polytechnics and other educational institutions to acquire university status and become new or post-1992 universities. In 2023/24, there were around 160 universities, just under three million students and a HE participation rate of just over 50% (HESA, 2025). Alongside the increase in the number of publicly funded universities, a system of private universities, boosted by the provisions in the Higher Education Act (2004), has taken hold and is growing.

How the marked expansion in student numbers would be funded was answered differently in different parts of the UK. In England, out went state funding and in came user charges in the form of capped tuition fees. The response in England reflected a view that the provision of Higher Education is best left to the market. In contrast, in Scotland, the principle of free access to HE, at least for Scottish students studying at Scottish universities, was viewed as sacrosanct and remains in place to this day, though it is under increasing pressure as universities compete with other claims on taxpayer funds. In England, a system of capped tuition fees on domestic students and uncapped tuition fees on overseas students set up an incentive structure which saw universities subsidising domestic students from the fees paid by overseas students. Despite the principle of no tuition fees being maintained in Scotland, the level of funding from the state meant that a similar dynamic was set in motion in Scotland. Only recently did it become apparent that the funding arrangements in England and Scotland were unsustainable. In response to broader political concerns, the UK government attached new conditions to UK student visas, which saw the number of overseas students decline and many universities put in an unsustainable financial position. Bolton and Lewis (2025) reported that the number of UK universities running a financial deficit increased from 21% in 20/21 to 40% in 23/24, with almost three-quarters (72%) expected to be in deficit by 25/26. The short-term response of the universities has been to axe courses and offload academic staff (Universities UK, 2025).

Massification, Marketisation and the Rise of Managerialism

In England, the HE system is well down the path to becoming fully privatised, with state control being replaced by the market mechanism (Marginson, 2017; Tomlinson & Watermeyer, 2022). The fact that governments of various hues have come and gone since the marketisation process got underway, and none have sought to reverse the changes, suggests that in England, leaving the provision of HE to the market is no longer a seriously contested position. Marketisation entails the commodification and monetisation of teaching and research. Deem (1998) was among the first to document the emergence of the managerialist culture, which was required to make the marketisation of the system possible. Managerialism manifested itself in a number of fundamental changes to how the HE system worked: students became consumers, management became obsessed with efficiency and gaining a "competitive edge" on their rivals

rather than valuing education for its own sake, trade unions were marginalised and Human Resources departments became an increasingly powerful branch of management, consultation was replaced by authoritarianism, and cooperation within and between universities was replaced by competition. The language of business has been adopted by those managing universities. Courses became "product" or "part of our offer" and were to be delivered by a resource with capacity (formally an academic) to a customer (formerly a student) in a way which is expected by managers (formerly university Deans/Chancellors) to contribute to the university's surpluses. Relationships between staff and students became transactional and bound by legal contracts, with unhappy customers having the opportunity and inclination to sue providers if they did not get what they wanted from their supplier. Graduates were to be "work ready", and the entrepreneurial skills inculcated in them by their Professors put at the disposal of the economy.

Like any other sellers in a competitive market, universities will want to portray what they are selling in the best possible light. The fact that six UK universities were told to take down potentially misleading adverts by the Advertising Standards Agency (BBC, 2017) indicates that, at times, they may have been at risk of misleading their consumers. One way of overcoming the information gap faced by potential students in the market for HE services is for impartial experts to assess the important but difficult to assess quality dimension of a university's teaching and research. If these assessments are accurate and made public, they have the potential to make the market function more effectively. If a university receives a more favourable assessment of its teaching and research than a rival, its higher score becomes a basis for differentiating its products from those of its competitors. To the extent HE customers value the quality dimension of teaching and research and are willing to pay for it, higher scoring institutions will have pricing power over their lower scoring counterparts. The resultant higher revenue streams will then allow the university to outbid its rival in the market for scarce factor inputs (e.g., academic labour). This sets up a "Matthew effect" where highly rated institutions become more highly rated.

In the study of markets, it is generally accepted that the stronger the competition in a market, the greater is the pressure on sellers to squeeze as much out of their inputs as they can get. The ways university managers have come up with to do this are many and varied. If there is more than one way to do something, then efficiency demands that the lowest cost way is used, even if that is to the detriment of quality. One of the main ways this efficiency imperative has revealed itself in universities is in the growing casualisation of the workforce and the emergence of a new precariat (Willson et al., 2026). In true Taylorist style, the complex task of providing an education is decomposed into its constituent elements, with these being sold to the lowest bidder in a race to the bottom. The remaining permanent staff find that no matter how much work they do, in the eyes of their line manager, it is never enough. Further pressure comes from the continuing process of review by students who are invited to pass judgment on their lecturers' performance at the end of their course, and by their line managers. In an interesting spin on the evaluation culture which now dominates the lives of academic staff in UK universities, Erickson et al. (2021) surveyed UK academics to find out what they thought about the senior managers running their institutions. The responses painted a bleak picture in which many respondents viewed senior university management as inept, capricious, and sometimes malicious.

The Evolution of Research Assessment Exercises in the UK

Periodic assessments of university teaching and research are conducted to assess whether the funds put into the system by the state are being used effectively or not. Though it might be expected that the results of teaching and research assessments should be equally important to universities, few would argue that the results of research assessment exercises are of much greater importance to students and staff at universities. Linking funding to the results of research assessment exercises means universities have a direct interest in scoring as highly as they can. In addition to having a direct effect on a university's finances, research assessments may also have an indirect effect through their impact on a university's reputational capital. Research scores are an important determinant of where a university ranks in the increasingly important league tables produced by various organisations. Higher rankings will make it easier to attract students and give universities more pricing power, which allows them to fill their course places with students paying higher fees. Better research assessment results will make a university a more attractive place to work for ambitious research-oriented academic staff, setting up a virtuous circle from which universities benefit.

The research selectivity exercise of 1986 was the response of the then Thatcher government to pressures on the public finances and the demand on the part of universities for greater transparency in the allocation of research funds. It introduced the idea of Units of Assessment (UOAs) - disciplinary areas - with the aim of channelling funds to where they would be best used, and was followed by a similar exercise in 1989. Tapper and Salter (2003) and Bence and Oppenheim (2005) give good accounts of the evolution of the system in the early days. After this, there were a series of Research Assessment Exercises (RAEs) in 1992, 1996, 2001, and 2008, which were followed by the introduction of the Research Excellence Frameworks (REF) in 2014 and 2021. Each iteration of the process introduced changes to how research quality was measured and how it was rewarded. Though the early exercises were far from egalitarian, they took a slightly less elitist approach and provided some support for lower quality submissions. Over time, however, the system became more selective, with only research of the highest quality being funded. Not only has how research quality is assessed and rewarded changed, but what is included in these exercises has been widened to include aspects of research previously excluded. In response to criticism that the RAE was too academic and focused too much on academic outputs, the REF framework of 2014 broadened what was assessed to include the impact of research and the quality of the research environment. Not surprisingly, there has been a heated and extensive debate on how changes to research assessment have affected those working in universities. In addition to monitoring what is done in universities and whether state funds are being used effectively or not, research assessment exercises have the potential to overcome the problems created by asymmetric information and opportunistic behaviour in university departments.

Bias in Internal and External Assessments of Research

Given the range of areas covered by academics in their research, university research managers may lack the disciplinary expertise required to gauge research quality. Even if managers have the skill set needed to judge research quality, there may be difficulties with acting on that judgement, e.g., telling a colleague, who may also be a friend, that the quality of their research is not up to standard, which must be awkward and, as such, may be avoided. Inter-personal

relationships may result in research leads making unjustifiably favourable judgements on the research quality of members of their preferred group, and unjustifiably negative comments on others. These sorts of internal research assessment biases are incompatible with the efficient internal allocation of research funds. As well as potential bias originating with those doing the assessing, there are likely to be problems emanating from those being assessed. Academic staff have incentives in the form of bonuses, promotion, and access to research funds to exploit any information advantage they may have over those assessing their research by overstating the quality of their research. As well as being useful to potential students when deciding what to study and where, independent, expert assessments can be of use to universities and their staff. To the extent that these exercises yield unbiased, accurate assessments of research quality, they have the potential to help university managers allocate their scarce research funds more efficiently. The results of these assessments can also perform a useful function for academic staff by helping them to identify their own research strengths and weaknesses. Of course, if the assessments produced by external research assessment exercises are themselves biased, then the effectiveness of any decisions made by university managers which hinge on their objectivity and accuracy is likely to be reduced. In addition to the effect biased external assessments of research quality have on the internal allocation of research funds, they will also have a direct effect on the allocation of funds to the universities themselves.

As with all change, there are winners and losers, with the former singing the praises of the intensification of research measurement and others maligning the change. Johnston and Reeves (2017) identified three broad strands of criticism. Willmott (2011) coined the phrase "journal list fetishism" to describe the extent to which some academics have internalised the rankings in journal quality guides. This, he argues, has changed not only how academics see themselves but, more importantly, how they relate to one another, with damaging effects on the ethos of UK universities. The second strand of criticism relates to the competition for research assessment funds, which has pitted universities against one another, with the potential benefits of cooperation being lost. The third strand of criticism acknowledges the potential benefits of research assessment but criticises how it has been done. Others, such as Biagioli et al. (2019), explain how the research assessment is open to and is gamed by researchers in a way that threatens to undermine their effectiveness. Spence (2019) argues that the whole research assessment process creates a veneer of vibrant research activity which does not stand up to scrutiny. Some question the very idea of commensurability, which underpins the whole RAE edifice, e.g., if a university scores a three in its submissions to the Physics UOA and to the Film and Theatre UOA, what, if anything, does that tell the university about the quality of the research in both fields?

Supporters of efforts to measure research quality more closely claim it has contributed to an increase in the quality-adjusted quantity of university research. Checchi et al. (2019) in a study of 31 countries found that the introduction of Performance-Based Funding Systems led to a short-term increase in the average quality (proxied by citations) and quantity of research produced by universities. They could, however, not find evidence of an increase in the number of excellent research papers (proxied by the number of papers in top research journals) produced. Interestingly, Checchi et al. (2020) in a study of two iterations of Italy's research assessment regime found that it had induced a degree of convergence both within and between universities. Two factors explained this convergence. One, an increase in the research produced by researchers between the two iterations of the assessment process. Two, the recruitment of new, more productive researchers to add to or replace their less productive

peers who had left between the two assessments. Banal-Estañol et al. (2023) carried out a natural experiment when they compared the research performance of UK universities (where PBFS had been introduced) with US universities (where PBFS had not been introduced) to isolate the impact of the introduction of PBFS on research in UK universities. They found that it did lead to higher levels of output and research quality in the UK, but not to higher levels of research productivity as such.

Peer Review and 'Place of Publication' as Indicators of Research Quality

Comparisons of the merits of alternative approaches to the quantification of research quality assume that a 'true' measure of research quality exists and all that needs to happen is that it be measured. If there is no 'true' measure, then research assessment is at best a waste of resources and at worst a dangerous diversion (Martin, 2011). Research quality can be assessed *directly or indirectly*. Direct assessment typically involves independent disciplinary experts (peers) scrutinising publications in their field of expertise with a view to producing an assessment of quality. Indirect assessment does not involve direct scrutiny of a paper but makes use of factors such as the place of publication to make inferences about its quality. To some extent, the distinction between peer review and place of publication is an artificial one, as prior to being included in a journal, a manuscript will have gone through a peer review process. Indeed, one of the factors it is claimed that distinguishes highly ranked journals from their lower rated counterparts is the rigour of the review process. If a paper published in a peer reviewed journal is then included in a submission to a research assessment exercise, it will go through a second round of peer review. If the two reviews result in the same (or similar) assessments of quality, then there may be an efficiency case for replacing peer review with Journal Quality Guide (JQG) scores.

Given the high cost of research assessment exercises, their continued use suggests that those who fund them are of the view that they provide useful information on the important but difficult to assess research trait that is quality. But they are not without their problems. Who decides who becomes a panel member? How easy is it to avoid conflicts of interest? How are the different panel members' assessments of the same paper reconciled? Do hawks or doves dominate? Who gives ground and whose views dominate? Resource constraints mean that the outcome of peer review turns on the judgment of a small number of potentially over-worked panel members, e.g., in Panel I in the 2008 RAE 54 sub-panel and 10 main panel members reviewed over 12,000 papers by over 3,000 researchers. If assessors are over-stretched, this may lead to errors in quality measurement.

In comparison with the panel members in research assessment exercises, the referees used by journals may have more time and a wider range of appropriate expertise. If this is the case, then their quality assessments may be more accurate than those emanating from peer review. However, the use of a paper's place of publication will only lead to accurate inferences about its quality if a journal's rank in a journal quality list truly reflects the relative quality of the research it publishes and if all the articles published in the journal have similar levels of quality. Unfortunately for those attempting to quantify research quality, there is plenty of evidence to suggest that neither of these is the case. Different journal quality guides are based on different methodologies and have different, sometimes very different, rankings of the same journal. The first question which must be answered is whether 'place of publication' is to be used to proxy research quality, which quality guide is the best guide of quality? In the UK, the

Association of Business Schools (ABS) journal quality guide (JQG) dominates other guides such as SCIMAGO. Indeed, Willmott (2011) suggested that the UK's Business School academics have internalized its journal rankings to such an extent that they affect how they think about themselves and colleagues. In addition to the problem of how to rank different journal rankings, there is much evidence that the quality of the research published in journals is not uniform. Papers may find their way into lower ranked journals than they might have achieved for a variety of reasons: authors may underestimate the probability of a paper being published in a highly ranked journal, either because they have underestimated the quality of the paper or overestimated the rigour of the refereeing process; if the time between submission and publication is inversely related to journal quality ranking, those that publish high quality articles in low quality journals may be trading off speed of publication against journal ranking. Whatever the reason, the upshot of these sorts of factors is that high quality research may well be published in low quality journals and vice versa. Rowlinson et al. (2015) sum up the issue when they state:

> Although much top-quality work was indeed published in what are generally regarded as leading journals, top-quality work could also be found in journals occupying a lower position in conventional rankings. Similarly, some of the work considered that had been published in so-called leading journals was thought to be of less than top quality. (Rowlinson et al., 2015)

The lack of consistency in terms of what is published in journals is not the only problem associated with using the place of publication to make inferences about research quality. Research not published as a journal article would be omitted if quality assessments were based solely on journal articles. An approach to quantifying research quality that ignores research published in formats such as books would clearly be flawed. Outright mistakes or poor decisions made by referees and/or editors may allow low-quality articles to appear in the highly ranked journals and may prevent high quality articles from appearing in highly ranked journals. In a famous experiment, Peters and Ceci (1982) resubmitted twelve papers to the journals which had previously accepted them but changed the author names and institutional affiliation to make them appear less prestigious. Only three of the twelve resubmitted were accepted on resubmission, with the other nine being rejected. If this is representative of how the process works more generally, the idea that such a process is impartial and consistent is clearly questionable. Though one of the potential benefits of peer review is that it may detect errors rooted in the journal review process, it must be borne in mind that peer review itself is not immune to the same sorts of human error.

Despite the problems with relying on place of publication as a quality guide, there is much evidence that JQGs do a good job of predicting the results of peer review exercises (e.g., Groot & García-Valderrama, 2006; Kelly et al., 2009). Kelly et al. (2009) go as far as to state that the JQG in large measure mirrors the quality judgements made by the RAE panel. Indeed, the high degree of positive correlation between the two measures has been used to vindicate the methodology underpinning the ABS JQG. It is also seen as providing a justification for the use of the ABS journal quality list by university managers in their management of research within their various institutions. Even though peer review and place of publication result in very similar estimates of research quality, Roberts (2003) reminds us that one is not a perfect substitute for the other. Despite the merits of the place of publication, in the UK, research is still assessed by peers. Given the human element in the system, it is inevitable that some have suggested the possibility of error or bias.

Is there Bias in Peer Reviews of Research Quality?

If the purpose of RAEs is to make the allocation of research funding more efficient, then the more it is impartial and applied consistently across units and over time, the better. In essence, peer review involves people making judgments on the quality of the research of other people, and as such, it involves an element of subjectivity. Concern over the impartiality of the peer review process has existed for as long as people have attempted to measure the quality of the research produced by other people. Bias, conscious or unconscious, would involve a publication being assessed as having a quality level higher or lower than merited by its objective quality, with the source of any bias being the panel members themselves. If a UOA panel is overly generous in its assessment of the quality of the submissions from members of its own discipline, this would constitute a type of bias best described as nepotism. If all UOA panels are nepotistic to the same extent, this will lead to grade inflation across all UOAs but no change in their relative positions. In contrast, if one panel is more nepotistic than the others, this inequality will kick-start a dynamic which will result in the research performance of the more nepotistic UOA pulling away from that of the others. The other possible type of bias would involve a UOA panel being unduly harsh in its assessment of the quality of the research submitted to it. This type of negative bias is typically referred to as discrimination and is the mirror image of nepotism.

Doyle et al. (1996) claimed that in the UK, universities located in England are favoured at the expense of those in other parts of the UK. In relation to bias in peer-review based research assessment exercises, Johnes et al. (1993) posed the question:

> To what extent can the considerable differences in research rating between cost centres [subject areas] …be attributed to corresponding differences in research quality and to what extent are these differences due to the whims and prejudices of subject panels?

Mingers et al. (2012) went further when they drew attention to what they saw as the *"...inherent subjectivity and openness to bias..."* of the peer review process. Taylor (2011) claimed he'd uncovered evidence of overgenerosity on the part of the E&E panel at the 2008 UK REF. He stated: *"…there is evidence that the E&E panel was more generous in its assessment of research quality than the B&M panel. This suggests an element of bias in the decisions reached by the E&E panel."* Following up on this study, Clerides et al. (2011) in their study of submissions to the E&E UOA in the 1996 and 2001 RAEs could find no evidence of bias based on university age, location, or size. Doyle et al. (1996) found that having a member on the assessment panel was worth 0.822 of a grade point. Coupé (2003) finds the same bias in cross-sectional analysis of the 1992 and 1996 RAEs, though the result goes away when intertemporal variation to better account for quality is allowed. McKay (2012) and Sharp and Coleman (2005) could find no association between panel membership and RAE ratings. As is often the case with these sorts of questions, the picture presented by the empirical work is mixed and inconclusive.

The Determinants of REF Scores Other than Place of Publication

The extensive empirical work in this area (e.g., Johnes et al., 1993; Kelly et al., 2009; Taylor & Francis, 1993; Taylor, 1995) has identified a number of factors associated with variation in RAE scores. The size of a research group appears to play a key role in determining RAE scores, with

an inverted-U (quadratic) relationship between the number of research active staff submitted and RAE score being evident in assorted studies. It has been argued that the presence of economies and diseconomies of scale in the research process, and the fact that larger research groupings are more difficult to ignore, may explain why larger groups tend to do better in research assessment exercises than their smaller rivals. The idea that causality may run in the other direction and that research assessment might determine research group size has also been suggested. The number of research students (both an input to and an output from research activity) is also positively correlated with RAE scores. External research income, particularly in the scientific disciplines, is positively correlated with research performance. We now discuss how decomposition techniques might be used to tease out and measure discrimination or bias in RAEs.

Decomposing REF Scores into Explained and Unexplained Components

The Blinder-Oaxaca decomposition (Blinder, 1973; Oaxaca, 1973) assumes that if all the factors that explain the explained variable are included as explanatory variables in the estimated equation, then that part of the difference in group averages which is not explained by all relevant explanatory variables is by definition unexplained and must be the result of bias or discrimination. Note that if variables which explain the difference in mean scores are not included in the explanatory model, this will bias the estimates of bias or discrimination. If it is assumed that the REF score (R) of the i^{th} university, which submitted to either UOA_E or UOA_B (where lower case E denotes the Economics & Econometrics UOA and B denotes the Business and Management Studies UOA), is determined by a vector of research characteristics (X_{Ei}, X_{Bi}), then linear REF equations for each UOA can be written as follows:

$$R_{Ei} = \beta_E X_{Ei} + \varepsilon_{Ei}, i = 1, \dots N_E \qquad (1)$$

$$R_{Bi} = \beta_B X_{Bi} + \varepsilon_{Bi}, i = 1, \dots N_B \qquad (2)$$

where ε is a random error term, and N is the sample size. If the error term in equations (1) and (2) conforms to the standard assumptions of the classical linear regression model, namely that they are mean zero and independently and identically distributed, then the fitted regression, estimated by least squares, must pass through the sample means. The REF scores (evaluated at the means, ($\bar{R}_E$ and $\bar{R}_B$)) for each of the two UOAs can be written as follows:

$$\bar{R}_E = b_E \bar{X}_E \qquad (3)$$

$$\bar{R}_B = b_B \bar{X}_B \qquad (4)$$

where b_E and b_B denote the least squares estimates of β_E and β_B, and $\bar{X}_E$ and$\bar{X}_B$ are the mean values of the vector of determinants of research quality for the E&E and BMS UOAs, respectively. Subtracting equation (4) from equation (3), the gap between the mean scores in the two UOAs can be written as:

$$\bar{R}_E - \bar{R}_B = b_E \bar{X}_E - b_B \bar{X}_B \quad (5)$$

Following the Blinder (1973) and Oaxaca (1973) method, if ($b_E \bar{X}_B$)is added to and subtracted from the right-hand side of the equals sign in equation (5), the resultant equation can be rearranged to give a new expression for the difference in mean scores such that:

$$\bar{R}_E - \bar{R}_B = (b_E - b_B)\bar{X}_B + (\bar{X}_E - \bar{X}_B)b_E \qquad (6)$$

The first term on the right-hand side of the equals sign in equation (6) picks out the part of the difference in the UOA means ($\bar{R}_E$ - $\bar{R}_B$) due to differences in the regression coefficients (b_E and b_B), i.e. the unexplained or discriminatory part of the gap. The second term on the right-hand side of the equals sign captures the part of the difference in mean REF scores that is the result of differences in the vector of research characteristics ($\bar{X}_E$ and $\bar{X}_B$) in both UOAs.

Figure 1 below illustrates how the Blinder-Oaxaca method can be used to decompose the difference in the UOA mean REF scores. Two regression lines showing REF GPA regressed in ABS JQG GPA for both UOAs are shown in the figure, with the E&E UOA line having a steeper slope and lower intercept than its BMS counterpart. If it is assumed both UOAs were assessed by the E&E panel, then this allows us to generate a counterfactual mean REF score for the BMS UOA of $\overline{R_B^E}$. This is what submissions to the BMS UOA would have scored if they had been assessed by the E&E panel as opposed to the BMS panel. The actual mean REF score for submissions to the BMS A is denoted asdenoted $\overline{R_B}$. The difference between what the BMS submission would have earned if it had been assessed by the E&E panel and what it actually earned when assessed by the BMS panel is given by ($\overline{R_B^E} - \overline{R_B}$). This is the part of the difference between the two mean REF scores which is not explained, i.e. it is the result of difference in the slope parameters in the separate UOA regressions. The difference between the E&E panel's assessment of the work submitted to the E&E panel and what the BMS material would have earned if it had been assessed by the E&E panel is given by ($\overline{R_E} - \overline{R_B^E}$). This is the explained part of the difference in E&E scores and BMS scores, which, i.e. it is the part of the difference which is the result of the higher quality of the E&E research.

If the slope of the coefficients is equal, i.e. there is no difference in how the BMS and E&E panels map any given research quality scores onto REF scores, then any gap in mean REF scores will be the result only of differences in research quality. However, if there is a difference in the coefficients, then this may lead to a difference in mean REF scores even where there is no difference in mean research quality.

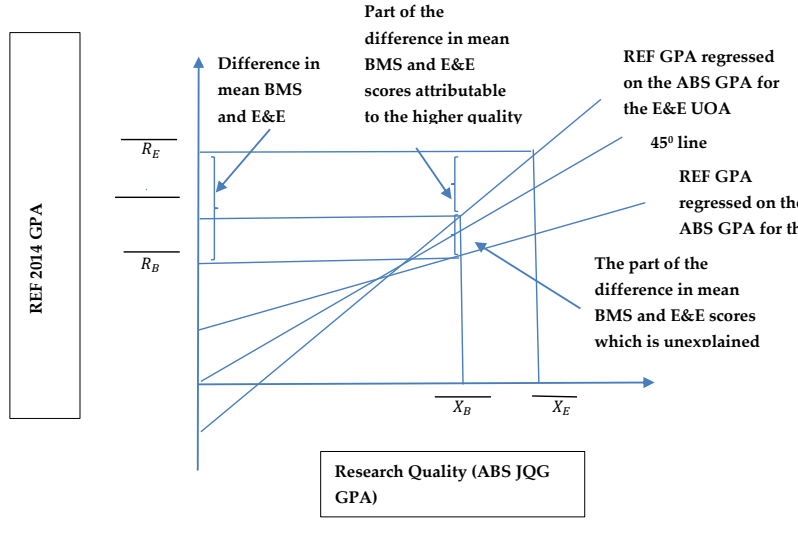

Figure 1. Decomposing the difference in mean REF scores in the E&E and BMS UOAs into its

explained and unexplained components

The fact that there are two measures of the unexplained part of the differential in group means is a well-known property of index numbers. In this context, the alternative decomposition would involve asking what would have happened if both submissions, rather than being evaluated by the E&E panel, had been evaluated by the BMS panel. This would generate a different counterfactual. The decomposition would be derived by adding and subtracting the term $(b_B\bar{X}_E)$ to the right hand-side of equation (5), as shown in equation (7):

$$\bar{R}_E - \bar{R}_B = (b_E - b_B)\bar{X}_E + (\bar{X}_E - \bar{X}_B)b_B \quad (7)$$

It should be noted that there is no a priori reason for preferring one approach to the other. One way to deal with this dilemma might be to use a weighted average of the two estimates. Though this approach makes use of both estimates, gives a single estimate and makes the indeterminate determinate, there is no a priori reason for preferring this to either of the other estimates. Daymont and Andrisani (1984) extend the Blinder (1973) and Oaxaca (1973) approach by adding interaction terms as shown in equation (8) below:

$$\bar{R}_E - \bar{R}_B = (b_E - b_B)\bar{X}_E + (\bar{X}_E - \bar{X}_B)b_B + (b_E - b_B)(\bar{X}_E - \bar{X}_B) \quad (8)$$

If the third term on the right-hand side of the equals sign in equation (8) is positively signed, then this indicates that submissions to E&E UOA tended to have higher values for the variables that were weighted more highly by the E&E panel. To try to put some empirical flesh on these empirical bones, a data set was created.

The Data

Two data sources were used in this study: one, the 2010 ABS JQG; two, the REF 2014 results, which are both available online. The sample of universities which were included in the study included all 28 universities which submitted to the E&E UOA and 28 of the 106 universities which submitted to the BMS UOA at REF2014. Over 10,000 papers were submitted to the BMS UOA at REF2014. LBS, for example, submitted 356 papers. To try to gather information on all of the submissions of all of the universities included in the BMS UOA would have been enormously time consuming and required resources not available to the researchers. As such, it was decided that a sample of 28 – over one quarter of all the universities which submitted to the BMS UOA - would be sufficient. It was also equal to the number of submissions to the E&E UOA, giving a balance to the study. It is worth pointing out that, as all of the universities which submitted to the E&E UOA were included in the sample, the parameter estimates are the population parameters and not sample-based estimates of the parameter values. Given the composition of a sample is as, if not more important, to the successful conduct of inferential statistics, an effort was made to create a BMS UOA sample which mirrored the underlying population in terms of factors such as the size and age of the universities included in the sample. Table A3 in the appendix shows the universities included in both samples. It includes 9 post-1992 universities, 8 of the elite, research intensive Russell Group universities, and 11 other types of universities, with the universities in the different regions and countries of the UK represented. The universities included in the sample were included to give a sample which was representative of all the universities which submitted to the BMS UOA. Interestingly, perusal of Table A3 reveals that not a single post-1992 university submitted to the E&E UOA. As the absence of a submission to the E&E UOA is often a precursor for the removal of an

undergraduate course in the subject (Johnston & Reeves, 2018), this is a cause for concern. The study analyses 56 submissions to both UOAs and attempts to explain their variation using a small number of key variables. It has to be said that decomposition techniques are normally deployed using very large data sets. Given the size of the underlying populations in this study, all submissions to the E&E and BMS UOAs, the numbers are never going to be large, and this has to be borne in mind as it may have implications for the stability and reliability of the estimates derived in the empirical work.

The 2010 ABS JQG was used as this was the one available at the time of REF 2014. It ranked journals on a scale of 1 to 4, 1 being the lowest quality and 4 the highest. The guide is revised every few years to reflect the fluctuating status of different journals. The language used to define the 1-4 categories is vague and allows a good deal of discretion about which categories journals should be placed in. For example, consider the issue of refereeing standards: papers submitted to both 4 and 3 rated journals that are not immediately rejected are 'heavily refereed', while for 2-rated journals they are 'fully refereed', and for 1-rated journals they are 'relatively lightly refereed'. Regarding originality, the 'most original' papers are said to be published in 4-rated journals, while 3-rated journals carry 'original and well executed' research. However, for 2 and 1-rated journals, there is no mention of originality. Given the vagueness of the definitions, positioning of journals on the 1-4 scale is ultimately a matter of judgement, and inevitably there will be disagreements about how some journals should be rated. Table A.2 in the appendix provides the definitions used in the ABS journal quality list.

The 2014 REF was a peer assessment of the quality of UK universities' research in all subjects. The REF was undertaken by the four UK higher education funding bodies, the Higher Education Funding Council for England (HEFCE), the Scottish Funding Council (SFC), the Higher Education Funding Council for Wales (HEFCW) and the Department for Employment and Learning, Northern Ireland (DEL). The REF results were used to allocate research funds to UK universities on the basis of quality from 2015-16. The REF was managed by the REF team, based at HEFCE, on behalf of the four UK higher education funding bodies. In all, 154 UK institutions submitted research in 36 units of assessment (UOAs). The submissions were assessed by panels of subject experts to produce an overall quality profile for each submission.

The overall quality profile awarded to each submission was derived from a sub-profile for each of the three elements of the assessment, weighted as follows:

1. The **quality of research** output makes up 65 per cent of the assessment.

2. The **impact** of the research makes up 20 per cent of the assessment.

3. The **research environment** makes up 15 per cent of the assessment.

The ABS JQG and the panels both use a 4-point scale to describe quality. Is it then possible to say that a paper in a 4 rated journal is equivalent to a 4* rating assigned to the paper by the REF panel? After all, it seems reasonable to assume the best on the ABS scale should correspond with the best on the REF panel scale. So, are the two scales commensurate? For the 4* category, the REF panel is looking for world-leading originality, significance and rigour. There is no mention of world-leading significance and rigour, nor, for that matter, anything about world-leading anything in the ABS Guide. Unlike the ABS JQG, the REF definitions of research quality categories say nothing about submission and acceptance rates, refereeing or

citation impact factors. The only thing the definitions have broadly in common is the word 'originality'. But not all papers in 4 rated journals make an original contribution to their subject; for example, review papers are a feature of many top journals. We can make similar comparisons down the scale and find similar issues, which make it very difficult to equate 4 and 4*, 3 and 3*, 2 and 2* and 1 and 1*.

Matching Papers Submitted to the REF to the ABS Journal Quality List

The REF results site provides details on each paper submitted by each university and for each UOA. The information includes the paper title, the journal in which it was published and the URL where the paper can be viewed. Viewing the paper online allows the author(s) to be identified. For each paper, the ABS JQG was used to determine its journal rating, and when the process was complete, a quality profile was produced by calculating the numbers in each category 1-4. This was a time-consuming task, especially when one considers the number of papers submitted by the larger universities. For example, in BMS, London Business School submitted 356 papers and LSE 296 papers; in E&E, Oxford submitted 242 papers and LSE 183 papers.

Descriptive Statistics

Table 1 below presents the mean and median values of the variables used in the regression and decomposition analysis. Rows 1 and 2 show that the quality of submissions to the E&E UOA, whether measured by REF or ABS JQG scores, was higher in the E&E UOA than in the BMS UOA. Row 3 shows that the difference (GAP) between the mean ABS and REF scores was higher in the BMS UOA than in E&E. Column 4 and rows 1 and 2 of Table 1 show that although the mean ABS score for the BMS UOA was 0.93 of the mean ABS score for the E&E UOA, the mean REF score for BMS was only 0.91 of its E&E counterpart. This small difference is consistent with the suggestion advanced by Taylor (2011) in relation to the 2008 RAE that the E&E UOA panel may have been a little more generous than the BMS panel when assessing research quality. However, it is also consistent with the possibility that the ABS JQG is more generous in its assessment of BMS journals than E&E journals, and that the REF process is simply correcting for this source of bias. Of course, these explanations are not mutually exclusive or exhaustive.

Table 1. Mean (Median) values of the variables used in the regression broken down by UOA[1]

Variable	Business and Management (BMS) UOA	Economics and Econometrics (E&E) UOA	BMS mean (median) as a proportion of the E&E mean (median)
ABS GPA	3.101 (3.038)	3.322 (3.287)	0.933 (0.924)
REF GPA	2.669 (2.595)	2.928 (2.986)	0.911 (0.869)
GAP	0.432 (0.425)	0.394 (0.362)	1.096 (1.174)
SUBJECT	1.000 (1)	0.000 (0)	Undefined (Undefined)

TOTINC5	£2,956,935 (£1,609,030.00)	£4,473,575 (£1,705,031.00)	0.660 (0.943)
BISRC	£732,617.00 (£231,132.50)	£2,368,795 (£732,366.00)	0.309 (0.316)
EUGB	£497,314.10 (£213,861.50)	£436,782.30 (£152,109.50)	1.138 (1.405)
UKIND	£454,945.20 (£94,143)	£253,795.90 (£192,394.90)	1.792 (0.489)
UKGOV	£699,349.90 (£585,687)	£545,689.10 (£266,147.60)	1.281 (2.200)
TOTDOC5	38.901 (27.4)	38.808 (28.50)	1.001 (0.961)
HCAC	32.857 (25)	28.500 (25)	1.150 (1)
FTEA	30.644 (23.1)	26.987 (23.975)	1.135 (0.963)
NEW	0.392857 (0)	0.0 (0)	Undefined (Undefined)
RG	0.285 (0)	0.571 (1)	0.499 (0)
ABS non-rated	15.067 (14.05)	16.957 (16.2)	0.888 (0.867)
REF unclassified	0.878 (0)	0.510 (0)	1.265 (0)
Number of Cases	28	28	-

1. All variable definitions, and their mean, minimum, maximum and median values can be found in Table A1 in the appendix.

The Distribution of ABS GPAs in Submissions to the BMS and E&E UOAs

Figure 2 below shows the distribution of the ABS JQG grade point averages (GPAs) for the universities which submitted to the BMS and E&E UOAs at REF2014. Table A1 in the appendix provides a detailed explanation of how the GPAs were calculated. Clearly, there was a higher proportion of universities with lower ABS GPAs in the BMS UOA than in the E&E UOA.

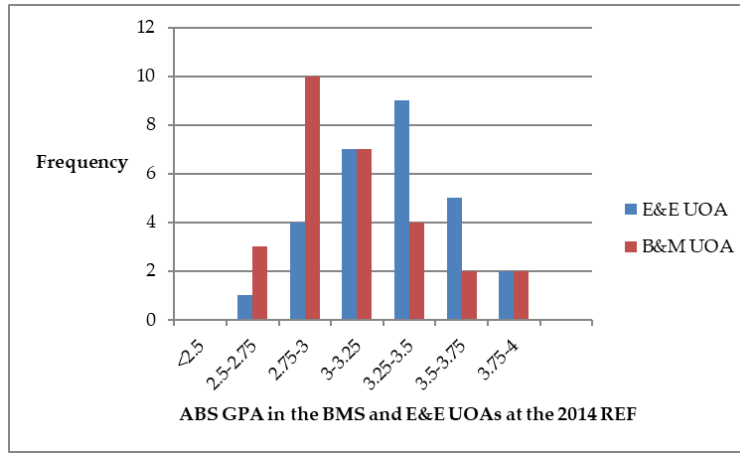

Figure 2. The distribution of ABSJQG GPAs in the BMS and E&E UOAs in the 2014REF

The Distribution of Average REF Scores in the Two UOAs

Similarly, and perhaps unsurprisingly, Figure 3 shows that a much higher proportion of REF scores in BMS were located at the lower end of the range of scores. In comparison to the BMS UOA, there are a higher number of E&E submissions in all categories with a GPA of at least three, and fewer in all categories below three.

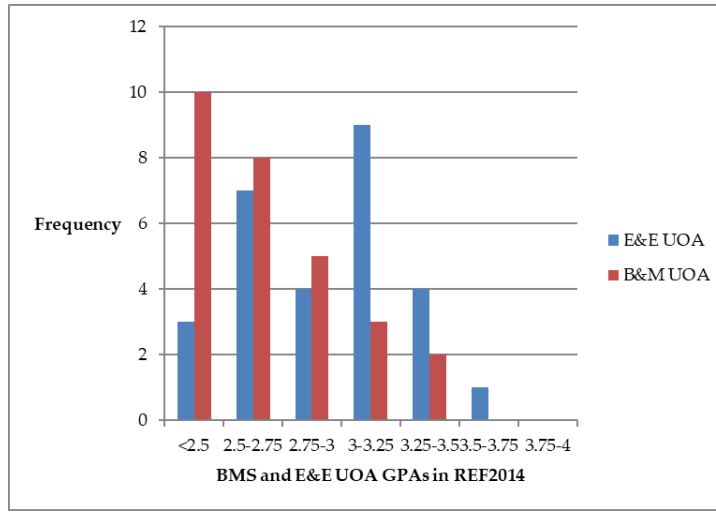

Figure 3. The distribution of REF2014 GPAs in the BMS and E&E UOAs in the 2014REF

Regressing REF Scores on ABS JQG Scores for the E&E and BMS UOAs

There is little in the way of theoretical priors on which to base predictions of differences in the slope and constants of the E&E and BMS regressions. The fitted equations could start from the same or different intercepts, be parallel, divergent or convergent over the range of ABS values found in the data. They might cross, but they might not. Table 2 below reports the results of least squares regressions of REF scores on research quality as proxied by the ABS JQG score and other explanatory variables. The regressions support the results of earlier research in that

in both UOAs, REF and ABS JQG scores are closely and positively correlated. It is also evident that the fitted E&E UOA equation has a steeper slope but starts from a lower intercept than its BMS counterpart. This indicates that any given increment in the ABS GPA is associated with a higher increment in REF scores for submissions to E&E than for submissions to BMS. The higher intercept for the BMS equation implies that BMS submissions at the lower end of the quality distribution were associated with higher REF scores than those in E&E. The negative intercept in equation II suggests that in the E&E UOA, a REF score of zero will result from an ABS score of 0.56. The fact that the E&E equation is steeper and starts from a lower intercept than the BMS equation implies that the equations are convergent at ABS scores lower than the value of the ABS score at which both equations meet. Solving regression equations, I (REF= 0.017 + 0.855ABS) and II (REF= -0.6 + 1.062ABS) yield a cross-over value for the ABS of 2.98. So, REF submissions to BMS with ABS JQG scores of less than 2.98 were associated with higher REF scores in BMS than in E&E. In contrast, submissions with scores of more than 2.98 in terms of the ABS JQG gained higher REF scores in E&E than in BMS. Perhaps the E&E panel was more severe on lower quality work than the BMS panel, or the BMS panel was less generous to higher quality work than the E&E panel.

Table 2. Regression analysis (Dependent variable is REF 2014 GPA)

Variable	**(I) BMS UOA**	**(II) E&E UOA**	**(III) Both UOAs With Subject Dummy**	**(IV) BMS UOA**	**(V) E&E UOA**	**(VI) Both UOAs**
Constant	0.017 (0.07)	-0.600 (1.65)	-0.205 (0.89)	-0.173 (0.38)	-1.259 (3.30) ***	-0.573(2.13) **
ABS GPA	0.855(10.15) ***	1.062 (9.700) ***	0.943(13.76) ***	0.813(4.70) ***	1.032 (8.83) ***	0.860 (9.17) ***
Economics Dummy[2]	-	-	-0.050 (1.12)	-	-	-
Log EUGB	-	-	-	- 0.0001(1.57)	-0.0099(1.59)	-0.0002(2.25) **
Log UKIND	-	-	-	0.00006(0.6 1)	0.0118 (1.93) *	0.00004 (0.4)
Log BISRC	-	-	-	0.025 (0.93)	0.0536 (2.00)	0.045(2.52) **

[2] Dummy equals 1 if submission to E&E UOA, 0 otherwise.
As no post-1992 universities submitted to the E&E UOA in 2014, the Economics subject dummy was not included in the pooled regression (column, VI).

					*	
HCAC	-	-	-	- 0.0002(0.15)	-0.001 (0.38)	-0005(0.44)
ABS non-rated	-	-	-	- 0.0016(0.59)	0.0052 (1.70)	0.0003(0.19)
REF Unclassified	-	-	-	0.0049(0.22)	-0.068 (2.57) **	-0.015(0.87)
Russell Group Dummy	-	-	-	0.033 (0.41)	0.070 (1.15)	0.049(1.00)
R-squared (adjusted)	0.79	0.775	0.803	0.791	0.853	0.833
No. of Cases	28	28	56	28	28	56

****, **, * indicate significance at the 1%, 5% and 10% levels respectively.*

The Role of Factors Other than Research Quality in Explaining REF Scores

The parameter on the economics dummy variable included in regression III of Table 2 was small in size and statistically insignificant, and suggests that after controlling for research quality differences, being in the E&E UOA did not shift the E&E regression up or down. Other equations (estimated but not reported) showed that being a post-1992 university had no independent effect on REF score, a finding which supports the results of previous research by Kelly et al. (2009). Similarly, the results show Russell group universities do better in REF exercises, not because they are members of the Russell group but because they do higher quality research. The R-squared values in equations IV and V indicate that the E&E REF scores are better explained than the BMS results, and that the inclusion of variables which capture the amount and source of a university's external funding added considerably to the explanatory power of the E&E UOA regression. In particular, higher levels of funding from UK Industry (UKIND) and Research Councils (BISRC) boosted REF scores and suggest a degree of complementarity between academic and commercial research. In contrast, higher levels of funding from the European Union (EU) seemed to be associated with lower REF scores. While the number of staff with research responsibility (HCAC) did not appear to have a statistically significant effect, the greater the proportion of research categorised as unclassified by the REF, the lower the REF score in E&E, a finding which is consistent with the heavy emphasis placed on academic journal articles in E&E. Interestingly, no such effect was evident in submissions to the BMS UOA. Having estimated regressions on the E&E and BMS

UOAs, we now turn to decomposing the difference in the mean scores between the groups into their explained and unexplained parts using a number of different decomposition methods.

Decomposing the Average REF Scores into their Explained and Unexplained Components

Column 1 (rows 2 and 3) of Table 3 below reports the results of the Blinder (1973) and Oaxaca (1973) decomposition of the difference in mean REF scores between the two UOAs. Using E&E as the reference group, the results indicate that the overwhelming majority (73%) of the difference in the mean REF scores was attributable to differences in research quality, with the remaining part of the difference due to differences between the UOAs in the regression parameters. Different regression coefficients mean that there were differences in how research quality was mapped onto REF scores by the two panels. Had submissions to both the BMS and E&E UOAs been assessed by the E&E panel (the reference group in this case), then the mean REF GPA for the BMS UOA would have increased by around 0.07 (0.27*(2.928-2.669)) to 2.74. Table 3 also reports the results of the Daymont and Andrisani (1984) two and threefold decompositions. In relation to the twofold decomposition, the BMS UOA is taken as the reference group, and the results suggest that if the E&E UOA submissions had been evaluated by the BMS panel, only 9% of the difference in mean UOA scores would have been the result of slope differences. As explained above, the threefold-version of the Daymont and Andrisani (1984) decomposition extends the twofold model by incorporating an interaction term designed to capture the impact of the combination of slope and quality differences on differences in mean UOA scores. If, for example, submissions to the E&E UOA had higher values for explanatory variables valued more highly by the E&E panel, then under this approach the interaction term would be positively signed and would boost the E&E UOA's mean score relative to that of the BMS UOA. The negative sign reported in row 3 of column 3 suggests that submissions to the BMS UOA had lower values in the explanatory variables most highly valued by the BMS panel, and that this may have had a large part to play in explaining the unexplained part of the difference in mean scores between the UOAs. While the Blinder (1973), Oaxaca (1973) and Daymont and Andrisani (1984) decompositions are based on separate regressions for each UOA, the Oaxaca and Ransom (1988, 1994, 1998) decomposition reported in the final column of Table 3 is based on a regression on pooled data. This approach, based on the regression VI, suggests the unexplained part of the difference is 7%. Even though the different decompositions provide different estimates of how much of the gap in mean scores is not explained by the explanatory variables, all approaches yield a positive number and are consistent with the suggestion that part of the difference in mean scores is not attributable to differences in the variables expected to determine the difference in mean scores.

Table 3. Decomposing the difference in the mean REF GPAs of submissions to the E&E and BMS UOAs at the 2014 REF

Based on the regression I in Table 2 above				
Decomposition Method	**Blinder-Oaxaca**	**Daymont & Andrisani**	**Daymont & Andrisani Three-Fold**	**Ransom-Oaxaca-Neumark**
Explained	73%	91%	91%	83%
Unexplained	27%	9%	27%	17%

Interaction Term	-	-	-18%	-
Decomposition based on regression VI in Table 2 above				
Explained	72%	99%	99%	93%
Unexplained	28%	1%	28%	7%
Interaction Term	-	-	-27%	-

Conclusion

The analysis in this paper yields several insights, some more novel than others. The paper corroborates the finding in other studies of a high degree of positive correlation between the ranking of the 'place of publication' in the ABS JQG and peer review carried out at the UK REF 2014 in the E&E and BMS UOAs. However, it shows that the relationship between submissions to the E&E and BMS UOAs is more nuanced than may have previously been suggested. Research at the lower end of the quality range (as proxied by the ABS GPA) and submitted to BMS achieved higher REF GPA scores than the same quality of research submitted to the E&E UOA. At the top end of the quality range, the situation is reversed. These findings are consistent with the notion of growing elitism in research into and the study of Economics in UK universities, highlighted by Johnston and Reeves (2014, 2018), with the relatively low scores given to lower quality research submitted to the E&E UOA contributing to a withdrawal of programmes in the UK's new universities and a growing dominance of elite institutions. The results of the various decompositions undertaken in the paper are consistent with the idea that part of the gap between the mean scores in submissions to the E& and BMS UOAs at REF 2104 is not explained by the explanatory variables included in the analysis. Estimates of the size of this unexplained component vary considerably depending on the specification of the relationship and the decomposition method deployed. Various possibilities might explain the existence of a gap: one, there is an element of bias at play; two, the model is mis-specified; three, the ABS JQG itself includes an element of bias in how it categorises journals in the two UOAs examined in the study.

Declarations

Authors' contributions: The first author is the only contributor to this article.

Competing interests: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Publisher's note: Educational Management and Policy remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Orcid ID

James Johnston https://orcid.org/0000-0002-6003-0936

References

Banal-Estañol, A., Jofre-Bonet, M., Iori, G., Maynou, L., Tumminello, M., & Vassallo, P. (2023).

Performance-based research funding: Evidence from the largest natural experiment worldwide. *Research Policy, 52*(6), 104780. https://doi.org/10.1016/j.respol.2023.104780

Bence, V., & Oppenheim, C. (2005). The Evolution of the UK's Research Assessment Exercise: Publications, Performance and Perceptions. *Journal of Educational Administration and History, 37*(2), 137-155. https://doi.org/10.1080/00220620500211189

Biagioli, M., Kenney, M., Martin, B. R., & Walsh, J. P. (2019). Academic misconduct, misrepresentation and gaming: A reassessment. *Research Policy, 48*(2), 401-413. https://doi.org/10.1016/j.respol.2018.10.025

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources, 8*(4), 436–455. https://doi.org/10.2307/144855

Bolton, P., & Lewis, J. (2025). *Universities grip financial crisis – but at what cost to the nation?* (Research Briefing CBP10037). House of Commons Library.

Checchi, D., Malgarini, M., & Sarlo, S. (2019). Do performance-based research funding systems affect research production and impact? *Higher Education Quarterly, 73*(1), 45-69. https://doi.org/10.1111/hequ.12185

Checchi, D., Mazzotta, I., Momigliano, S., & Olivanti, F. (2020). Convergence or polarisation? The impact of research assessment exercises in the Italian case. *Scientometrics, 124*(2), 1439-1455. https://doi.org/10.1007/s11192-020-03517-2

Clerides, S., Pashardes, P., & Polycarpou, A. (2011). Peer review vs metric-based assessment: Testing for bias in the RAE rating of UK economics departments. *Economica, 78*(311), 565–583. https://doi.org/10.1111/j.1468-0335.2009.00837.x

Coupé, T. (2003). Revealed performances: Worldwide rankings of economists and economics departments, 1990–2000. *Journal of the European Economic Association, 1*(6), 1309–1345. https://doi.org/10.1162/154247603322752557

Daymont, T. N., & Andrisani, P. J. (1984). Job Preferences, College Major, and the Gender Gap in Earnings. *The Journal of Human Resources, 19*(3), 408. https://doi.org/10.2307/145880

Deem, R. (1998). 'New managerialism' and higher education: The management of performances and cultures in universities in the United Kingdom. *International Studies in Sociology of Education, 8*(1), 47-70. https://doi.org/10.1080/0962021980020014

Doyle, J. R., Arthurs, A. J., Green, R. H., McAulay, L., Pitt, M. R., Bottomley, P. A., & Evans, W. (1996). The judge, the model of the judge, and the model of the judged as judge: Analyses of the UK 1992 research assessment exercise data for business and management studies. *Omega, 24*(1), 13–28. https://doi.org/10.1016/0305-0483(95)00044-5

Erickson, M., Hanna, P., & Walker, C. (2021). The UK higher education senior management survey: A statactivist response to managerialist governance. *Studies in Higher Education, 46*(11), 2134–2151. https://doi.org/10.1080/03075079.2020.1712693

Groot, T., & García-Valderrama, T. (2006). Research quality and efficiency. *Research Policy, 35*(9), 1362-1376. https://doi.org/10.1016/j.respol.2006.07.002

HESA. (2025). *Higher education student statistics: UK, 2023/24 – Student numbers and characteristics* (Statistical Bulletin SB271). Higher Education Statistics Agency.

Johnes, J., Taylor, J., & Francis, B. (1993). The research performance of UK universities: A statistical

analysis of the results of the 1989 Research Selectivity Exercise. *Journal of the Royal Statistical Society, Series A*, *156*(2), 271–286. https://doi.org/10.2307/2982732

Johnston, J., Reeves, A., & Talbot, S. (2014). Has economics become an elite subject for elite UK universities?. *Oxford Review of Education*, *40*(5), 590-609. https://doi.org/10.1080/03054985.2014.959912

Johnston, J., & Reeves, A. (2017). Assessing research performance in UK universities using the case of the economics and econometrics unit of assessment in the 1992–2014 research evaluation exercises. *Research Evaluation*, *26*(1), 28–40. https://doi.org/10.1093/reseval/rvw021

Johnston, J., & Reeves, A. (2018). An investigation into the role played by research assessment in the socio-geographic fragmentation of undergraduate economics education in the UK. *Higher Education*, *76*(4), 589-614. https://doi.org/10.1007/s10734-017-0227-9

Kelly, A., Morris, H., & Harvey, C. (2009). *Modelling the outcome of the UK business and management studies RAE 2008 with reference to the ABS journal quality guide*. Unpublished manuscript.

Marginson, S. (2017). Elite, mass, and high-participation higher education. In J. C. Shin & P. Teixeira (Eds.), *Encyclopedia of international higher education systems and institutions*. Springer. https://doi.org/10.1007/978-94-017-9553-1_50-1

Martin, B. R. (2011). The Research Excellence Framework and the 'impact agenda': Are we creating a Frankenstein monster? *Research Evaluation*, *20*(3), 247–254. https://doi.org/10.3152/095820211x13118583635693

McKay, S. (2012). Social policy excellence – peer review or metrics? Analyzing the 2008 Research Assessment Exercise in social work and social policy and administration. *Social Policy & Administration*, *46*(5), 526–543. https://doi.org/10.1111/j.1467-9515.2011.00824.x

Mingers, J., Watson, K., & Scaparra, M. P. (2012). Estimating Business and Management journal quality from the 2008 Research Assessment Exercise in the UK. *Information Processing & Management*, *48*(6), 1078-1093. https://doi.org/10.1016/j.ipm.2012.01.008

Neumark, D. (1988). Employers' discriminatory behavior and the estimation of wage discrimination. *Journal of Human Resources*, *23*(3), 279–295. https://doi.org/10.2307/145830

Oaxaca, R. (1973). Male-female wage differentials in urban labour markets. *International Economic Review*, *14*(3), 693–709. https://doi.org/10.2307/2525981

Oaxaca, R. L., & Ransom, M. R. (1988). Searching for the effect of unionism on the wages of union and nonunion workers. *Journal of Labor Research*, *9*(2), 139-148. https://doi.org/10.1007/bf02685237

Oaxaca, R. L., & Ransom, M. R. (1994). On discrimination and the decomposition of wage differentials. *Journal of Econometrics*, *61*(1), 5-21. https://doi.org/10.1016/0304-4076(94)90074-4

Oaxaca, R. L., & Ransom, M. (1998). Calculation of approximate variances for wage decomposition differentials. *Journal of Economic and Social Measurement*, *24*(1), 55-61. https://doi.org/10.3233/jem-1998-0143

Oaxaca, R. L., & Ransom, M. R. (1994). On discrimination and the decomposition of wage differentials. *Journal of Econometrics*, *61*(1), 5-21.

Peters, D. P., & Ceci, S. J. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences*, *5*(2), 187–195. https://doi.org/10.1017/s0140525x00011183

Robbins, L. (1963). *Higher education: Report of the Committee on Higher Education*. HMSO.

Roberts, G. (2003). *Review of research assessment: A report to the funding bodies* (Report 2003/22). HEFCE.

Rowlinson, M., Harvey, C., Kelly, A., Morris, H., & Todeva, E. (2015). Accounting for research quality: Research audits and the journal rankings debate. *Critical Perspectives on Accounting*, *26*, 2–22. https://doi.org/10.1016/j.cpa.2013.05.012

Sharp, S., & Coleman, S. (2005). Ratings in the Research Assessment Exercise 2001: The patterns of university status and panel membership. *Higher Education Quarterly*, *59*(2), 153–171. https://doi.org/10.1111/j.1468-2273.2005.00288.x

Spence, C. (2019). 'Judgement' versus 'metrics' in higher education management. *Higher Education*, *77*(5), 761–775. https://doi.org/10.1007/s10734-018-0300-z

Tapper, T., & Salter, B. (2003). Interpreting the process of change in higher education: The case of the research assessment exercises. *Higher Education Quarterly*, *57*(1), 4-23. https://doi.org/10.1111/1468-2273.00232

Taylor, J. (1995). A statistical analysis of the 1992 Research Assessment Exercise. *Journal of the Royal Statistical Society, Series A*, *158*(2), 241–261. https://doi.org/10.2307/2983291

Taylor, J. (2011). The assessment of research quality in UK universities: Peer review or metrics. *British Journal of Management*, *22*(2), 202–217. https://doi.org/10.1111/j.1467-8551.2010.00722.x

Tomlinson, M., & Watermeyer, R. (2022). When masses meet markets: Credentialism and commodification in twenty-first century higher education. *Discourse: Studies in the Cultural Politics of Education*, *43*(2), 173–187. https://doi.org/10.1080/01596306.2020.1814996

Trow, M. (1973). *Problems in the transition from elite to mass higher education*. Carnegie Commission on Higher Education.

Universities UK. (2025). *Universities grip financial crisis – but at what cost to the nation?* https://www.universitiesuk.ac.uk/what-we-do/creating-voice-our-members/media-releases/universities-grip-financial-crisis-what

Willmott, H. (2011). Journal list fetishism and the perversion of scholarship: Reactivity and the ABS list. *Organization*, *18*(4), 429–442. https://doi.org/10.1177/1350508411403532

Willson, R., Stewart-Robertson, O., Julien, H., & Given, L. M. (2026). Academic casualisation and precarity: a scoping review. *Higher Education Research & Development, 45*(4), 1165-1187. https://doi.org/10.1080/07294360.2025.2573983

Appendix

Table A1

Table A1
Regression variable definitions and descriptive statistics

Variable Name	Description	Mean	Median	Min	Max
REF GPA	A weighted average was	2.798	2.7505	2.136	3.676

	calculated for all Universities that submitted to the E&E and BMS UOAs at the 2014 REF. The weights applied are the proportions of assessed outputs judged by the appropriate panel to have fallen into each of the five categories: UC (given a value 0), 1*, 2*, 3* and 4*.				
ABS GPA	The journal of each of the papers submitted by each institution to the 2014 REF in the BMS and E&E UOAs was identified and its ABS journal quality guide (2010) score noted. A weighted average ABS score for each submission by each university was calculated. Publications (journals, books etc.) not included in the ABS journal guide were excluded from the calculation. The weights applied in the weighted average calculations were the proportions of assessed outputs that fell into each of the four categories: 1*, 2*, 3* and 4*.	3.212	3.170	2.5213	3.855
GAP	This is simply a university's weighted average ABS score minus its weighted average REF score.	0.413	0.408	0.019	0.710
TOTINC5	Total Research Income & Research Income-In-Kind in £s received in the five-year period prior to the 2014 REF, i.e. from 2008-2013. It includes funding from UK Research Councils, Industry and	£3,715,255.00	£1,609,030.00	£298,000.00	£25,926,516.00

	Commerce, Central and Local government, and charities. It also captures funding from EU Government Bodies and other sources.				
BISRC5	The total amount in £s received in funding from BIS Research Councils, Royal Society, British Academy and Royal Society of Edinburgh in the five-year period leading up to REF2014.	£1,550,706	£483,904.5	£8,881	£24,196,075.00
EUGB5	The total amount in £s received from EU Government bodies in the five-year period leading up to the REE2014.	£467,048.20	£152,109.5	(£356)	£3349189.00
UKIND	The total amount in £s received from UK industry, commerce and public corporations in the five-year period leading up to the RFE2014.	£354,370.50	£27,434	(£92)	7188512.00
UKGOV5	The total amount in £s received from UK central government bodies, local authorities, health and hospital authorities in the five-year period leading up to the RFE2014.	£622,519.50	£254,764	£0	£6,715,827.00
TOTDOC5	Number of research doctoral degrees awarded by the University in the five years running up to REF 2014.	38.855	28.65	8	120.96
HCAC	Category A and C staff headcount.	30.678	25	7	101
FTEA	Category A staff Full Time Equivalent	28.816	23.625	6.8	98.83
NEW	Dummy =1 if the University received its charter in 1992 or after, 0 otherwise.	0.196	0	0	1

RG	Dummy =1 if the University a member of the Russell Group, 0 otherwise.	0.428	0	0	1
SUBJECT	Dummy =1 if submission is to the BMS UOA, and 0 if to the E&E UOA.	0.500	0.500	0	1
REF Unclassified	% age of a submission not classified as 1*, 2*, 3* or 4* by the REF panel.	0.694	0	0	4.9
ABS not rated	% age of submitted outputs published in journals not in the ABS journal quality guide or published in non-journal format.	16.012	15.95	1.9	53.6

Table A2

Table A2
REF2014 and ABS Journal Quality Guide (2010) assessment criteria

REF2014		ABS journal quality guide 2010	
4*	Quality that is world-leading in terms of originality, significance and rigour	4	All journals rated 4, whether included in the Journal of Distinction category or not, publish the most original and best-executed research. As top journals in their field, these journals typically have high submission and low acceptance rates. Papers are heavily refereed. Top journals generally have the highest citation impact factors within their field
3*	Quality that is internationally excellent in terms of originality, significance and rigour but which falls short of the highest standards of excellence.	3	3 rated journals publish original and well executed research papers and are highly regarded. These journals typically have good submission rates and are very selective in what they publish. Papers are heavily refereed. Highly regarded journals generally have good to excellent journal metrics relative to others in their field, although at present not all journals in this category carry a citation impact factor.
2*	Quality that is recognised internationally in terms of originality, significance and rigour.	2	Journals in this category publish original research of an acceptable standard. A well-regarded journal in its field, papers are fully refereed according to accepted

			standards and conventions. Citation impact factors are somewhat more modest in certain cases. Many excellent practitioner-oriented articles are published in 2-rated journals
1*	Quality that is recognised nationally in terms of originality, significance and rigour.	1	These journals, in general, publish research of a recognised, but more modest standard in their field. Papers are in many instances refereed relatively lightly according to accepted conventions. Few journals in this category carry a citation impact factor.
Unclassified	Quality that falls below the standard of nationally recognised work. Or work which does not meet the published definition of research for the purposes of this assessment.	Zero rated	This is the term used to refer to research either published in a journal not included in the ABS journal quality guide or in some non-journal format.

Table A3

Table A3
Universities included in the samples

BMS UOA Universities (n=28)	E&E UOA Universities (n=28)
Ulster	Birkbeck, London
Aberystwyth	Birmingham
York	Bristol
Bangor	Brunel
Dundee	Cambridge
King's London	City, London
Cambridge	UEA
Keele	Essex
Bristol	Exeter
East Anglia	Kent
Central Lancashire	Leicester
Bradford	UCL
Brighton	LSE
Anglia Ruskin	Manchester
Swansea	Nottingham
UWS	Oxford

Oxford Brookes	QMU London
SOAS	Royal Holloway London
Oxford	Sheffield
Nottingham Trent	Southampton
Sheffield Hallam	Surrey
London Business School	Sussex
Coventry	Warwick
LSE	York
Westminster	Aberdeen
Portsmouth	Edinburgh
Manchester Metropolitan	Glasgow
Southampton	St. Andrews