

Business and Management Studies in the United Kingdom's 2021 Research Excellence Framework: Implications for Research Quality Assessment

Blackburn, R., Dibb, S. & Tonks, I

Published PDF deposited in Coventry University's Repository

Original citation:

Blackburn, R, Dibb, S & Tonks, I 2023, 'Business and Management Studies in the United Kingdom's 2021 Research Excellence Framework: Implications for Research Quality Assessment', *British Journal of Management*, vol. (In-Press), pp. (In-Press).

<https://doi.org/10.1111/1467-8551.12721>

DOI 10.1111/1467-8551.12721

ISSN 1045-3172

ESSN 1467-8551

Publisher: Wiley

© 2023 The Authors. *British Journal of Management* published by John Wiley & Sons Ltd on behalf of British Academy of Management.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Business and Management Studies in the United Kingdom's 2021 Research Excellence Framework: Implications for Research Quality Assessment

Robert Blackburn ¹, Sally Dibb ² and Ian Tonks ³

¹University of Liverpool Management School, Liverpool, L69 7ZH, UK, ²Coventry University Business School, Coventry, CV1 5FB, UK, and ³University of Bristol Business School, Bristol, BS8 1PQ, UK

Corresponding author email: robert.blackburn@liverpool.ac.uk

Evaluating research is an established part of the research process, as funding agencies and governments seek to raise its quality and performance. The United Kingdom's Research Excellence Framework 2021 (REF2021) was the eighth formal assessment of research in UK universities. In Business and Management Studies (B&M), Sub-Panel 17, 108 universities submitted 16,038 research outputs and 539 impact case studies covering the period 2014–2020. Submissions were assessed by a panel of academic researchers and research users, nominated by a range of academic constituencies. The outcome was that the quality of UK research in B&M continues to improve since REF2014. The quality profile for REF2021 had 79% of research assessed as 3* (internationally excellent) and 4* (world-leading). The paper explains and reports on our experiences of the peer review process, analyses the outcomes and discusses the state of research within the discipline. Subsequently, we consider the wider implications of the REF process, its methodologies and impacts, contributing to the debate about research quality in universities. The paper concludes with support for peer review and expresses caution against the automation of research quality assessment.

Introduction

This paper aims to stimulate discussion and debate amongst the research community in relation to how research quality is assessed, particularly in business and management studies (B&M). In doing so, the paper charts the purpose, pro-

This paper draws upon the work of the United Kingdom's Business and Management Studies Sub-Panel 17 of the Research Excellence Framework (REF2021), and the authors would like to thank sub-panel members for their comments on an earlier draft of this paper. We would also like to acknowledge the helpful feedback from the anonymous referees, particularly in relation to broadening the scope and discussion of the paper. The views expressed are those of the authors.

cesses and outcomes of the B&M sub-panel in the United Kingdom's Research Excellence Framework (REF2021), providing an analysis of the results and a discussion of their wider ramifications. Performance measurement of university research is undertaken in many countries at periodic intervals using a range of performance-based research funding systems (PRFs) (Guena and Martin, 2003). Sivertsen (2017) notes that, as an early adopter of research evaluations, the best practice developed in the United Kingdom has affected the design and development of evaluation approaches employed in other countries.

A distinction is sometimes made between evaluation exercises that focus on research quality and/or those that also consider the consequences

for research funding allocation (Hicks, 2012). In this paper, we focus on research evaluation issues, presenting an analysis of REF2021 and sharing our relevant experiences of the B&M sub-panel.¹ The principles and approaches of the REF are outlined, followed by a discussion of the processes and outcomes of B&M Sub-Panel 17 (SP17). We recognize that the REF stimulates discussion and debate, spanning detailed attention on the rules, processes and results, through to more existential considerations about the future of quality assessments and their implications for future research activities (see e.g. FRAP, 2022; Geuna and Piolatto, 2016).

Background

UK universities and their academic disciplines have been subject to periodic reviews of research quality by governments since 1986. REF2021 was organized under the auspices of Research England, part of UK Research and Innovation (UKRI),² on behalf of the four UK higher education funding bodies. The REF aims to robustly assess the quality of UK university research across all academic disciplines, with the purpose of ensuring accountability and demonstrating the benefits achieved from the investment of public funds in research. The results of the assessment exercise are of considerable importance to universities, as they determine the funding allocation each receives to support their future research activities and are a marker of their reputation. The outcomes are also reflected in a range of university quality rankings and league tables, which are influential in the higher education market both within the United Kingdom and internationally.

Evaluations of university research can be regarded as an application of the new public management and the related themes of public accountability and performance evaluation (Bessant *et al.*, 2003; Franco-Santos and Otley, 2018; Hood,

1995). Cooper and Otley (1998) and Pidd and Broadbent (2015) provide analyses of previous exercises and their outcomes in relation to business and management research. The most recent assessment of UK research performance for the period 2014–2020, REF2021 published in May 2022, assessed 79% of the overall quality of B&M research as 3* (internationally excellent) and 4* (world-leading) (REF, 2022a). These B&M results provide evidence that the research evaluation system in the United Kingdom has been effective in improving research performance by ‘embedding a culture of research management across all HEIs’ (De Boer *et al.*, 2015, p. 114).

Following changes to the REF2014 rules, REF2021 is considered to give a more comprehensive view of UK research activity and quality than in previous assessments. The overarching principles and methodologies for REF2021 were, however, broadly the same as those applied in REF2014. Universities received separate ‘quality profiles’ for each discipline-based unit of assessment submitted, as well as a profile for the institution as a whole. Each quality profile was based on three separate sub-profiles, relating to outputs, impact and the research environment, as summarized in Table 1.

The overall profiles are weighted by the number of full-time equivalent staff (FTE) to account for submission size. This provides a picture of the quality of submitted research activity by institution and by discipline (i.e. for each of the 34 REF units of assessment), as well as for the UK higher education sector as a whole.

The UK REF is considered to have led the way for PRFs globally, although many other nations have adopted such approaches to evaluate publicly funded research and/or to allocate research funding. For example, a 2016 review of European Union Member States revealed that most had carried out PRFs (Jonkers and Zacharewicz, 2016),³ albeit using a variety of approaches reflecting national administrative traditions (Bleiklie and

¹B&M was one of 34 discipline-based units of assessment (UoA) in REF2021, with submissions to each UoA assessed by a sub-panel of academic researchers and research users (<https://www.ref.ac.uk/>).

²UKRI is a non-departmental public body of the UK government that oversees funding for research and innovation, on behalf of the Department for Business, Energy and Industrial Strategy (BEIS), which replaced the Department for Business, Innovation & Skills (BIS) in July 2016.

³In arguing for the efficient allocation of funding for research, the European Commission (in 2012) called for Member States to ‘introduce or enhance competitive funding through calls for proposals and organisational assessments as the main modes of allocating public funds to research and innovation, introducing legislative reforms if necessary’. These approaches aimed to stimulate research productivity, in terms of its volume, quality and socio-economic impact (Geuna and Martin, 2003; Hicks, 2012).

Table 1. Descriptor and weights for each element of assessment

Element of assessment	Description	Weight in overall profile
Research outputs	Institutions submitted an average of 2.5 outputs per FTE. Each individual output was assessed and graded	60% (reduced from 65% in 2014)
Research impact	Institutions submitted impact case studies (ICS), each of which was assessed and graded separately. The number of ICS submitted was determined by the size of the institution's submission	25% (up from 20% in 2014)
Research environment	A statement of the environment relating to B&M research was assessed and graded	15%

Michelsen, 2015).^{4,5} The UK system's reliance on three elements, outputs, impact and environment, is unusual compared with other PRFs which, more typically, focus on the assessment of research publications. It may be argued that the combination of these three elements, and the peer review processes of REF2021, have had a positive effect on research quality and its composition in UK higher education institutions (HEIs). The REF requirements for submitting institutions to demonstrate clear and appropriate research strategies, following the three elements, has undoubtedly influenced the strategic direction, internal processes and investment priorities of universities, but this has come at a cost. However, evidence suggesting that the scores of these three elements are highly correlated has generated debate about whether the inclusion of all three elements unnecessarily adds to the cost and complexity of the evaluation exercise (Pinar and Horne, 2022).

Rule changes for REF2021

Following REF2014, the Department of Business, Innovation and Skills published a Green Paper

⁴Important points of difference between the PRFs included when the system was implemented, the proportion of research funding allocated, which stakeholders were involved and the assessment approaches and criteria used (see Jonkers and Zacharewicz, 2016). Among the measures used were PhD student numbers, amount of competitive research funding, level of internationalization and scientific productivity (the latter typically assessing the number and quality of research publications, and measured using quantitative/bibliometric approaches, peer review or both; Zacharewicz *et al.*, 2019).

⁵For more information, see recent reviews of PRFs by Dougherty *et al.* (2016), Hicks *et al.* (2015), Pinar and Horne (2022), Sivertsen (2017) and Zacharewicz *et al.* (2019).

(BIS, 2015) consulting on the future shape of the higher education sector. This announced that the next REF would take place by 2021 (paragraph 4, p. 72). Notwithstanding the desire for a more cost-effective process using research metrics, the publication confirmed the importance of the Haldane Principle and peer review for research funding (paragraph 4, p. 69). In the event, the Wilsdon (2015) report on research metrics further confirmed the importance and desire for peer review, although this has not eliminated advocates of their use for assessments (e.g. Basso and di Tollo, 2022).

One of the main uses of the data from assessment exercises is to measure performance over time. Hence, the BIS report also commented on the improving quality of university research, noting that between the two periods 2001–2007 and 2008–2013, across all disciplines, 'The results of REF2014 showed significant improvements in the quality of research produced by UK universities since 2008: with 72% of UK universities' research outputs now assessed at the highest international levels compared with 51% in 2008' (BIS, 2015, paragraph 2, p. 70).

Within B&M, Pidd and Broadbent (2015) reported on the substantial increase in 4* outputs relative to RAE2008, which increased to 20.5% (up from 13.8% in Business and Management and 5.9% in Accounting and Finance, respectively).⁶ The authors suggested a range of factors to explain this improvement, including '... better and more ambitious research, an increased emphasis on internationalisation, greater receptiveness of international journals to European research and

⁶The Research Assessment Exercise (RAE) was the predecessor of the REF. In the 2008 RAE, Business and Management was a separate unit of assessment from Accounting and Finance, but the latter was subsumed into Business and Management from the REF 2014 onwards.

Table 2. Summary of main changes to REF2021 rules

Changes made	Description
All staff with 'significant responsibility for research' to be submitted	<ul style="list-style-type: none"> Clearly defined criteria applied Linked to Stern's no-selectivity principle^a
Required number of outputs to be submitted per FTE member of staff	<ul style="list-style-type: none"> Calculated as the total number of FTE members of staff multiplied by 2.5 Reduced from four outputs per staff member in 2014
Decoupling of staff and outputs	<ul style="list-style-type: none"> Allowed greater flexibility in the selection of outputs for return between a minimum of one and a maximum of five for each staff member Output pool could include the outputs of former staff, leading to an increase in number of duplicate outputs submitted by the same researcher who had moved institutions^b
Changes made to the percentages allocated to each of the sub-profiles	<ul style="list-style-type: none"> Increase in the contribution of the impact element of the assessment profile to 25% (from 20% in 2014) Reduction in the output element to 60% (from 65% in 2014) The environment stayed the same at 15%.

^a Universities were at liberty to employ staff on a variety of contracts with no requirement to undertake research, and thus avoid their requirement to be submitted to REF. They had to submit a Code of Practice document (CoP) that explained how the institution defined who was an independent researcher with 'significant responsibility for research'. The CoPs were submitted for scrutiny and approval to the REF Equality and Diversity Advisory Panel (EDAP), then published in time for the submission deadline (<https://www.ref.ac.uk/about-the-ref/blogs/defining-significant-responsibility-for-research-an-inclusive-approach/>).

^b This was linked to Stern's no-portability principle, which was amended as a transitional arrangement for REF2021.

increased selectivity in submissions when compared to RAE 2008' (Pidd and Broadbent, 2015, p. 574). Concern about the extent of selectivity across disciplines in REF2014 was revealed in data from the Higher Education Funding Council for England (HEFCE), which showed that 43% of eligible B&M staff were submitted, with a clear gender bias: 34% of women included compared with 48% of men (HEFCE, 2015). Having announced the timeline for REF2021, the Department of Business, Energy, and Industrial Strategy (BEIS) commissioned the independent Stern Report to review the rules for the research assessment process. This report (BEIS, 2016) proposed changes to the rules governing future submissions. The report addressed how institutions in the previous REF had been selective over staff included and noted the practice of 'buying-in outputs' by recruiting those with high-quality publications later in the REF cycle. Hence, the two principles of 'no selectivity' of staff and 'no portability' of outputs were proposed to address these behaviours and to achieve a more accurate view of the scale and quality of research in UK universities. 'No selectivity' meant that universities should submit all research-active staff and 'no portability' referred to the principle that outputs should be submitted by the institution at which the output had been

produced. In previous research assessments, outputs moved with the researcher and the institution where the member of staff was employed on the census date could submit the outputs.⁷

The Stern Report stimulated a swathe of opinion and debate amongst the research community (BEIS, 2016). A consultation with the sector found support for the no-selectivity principle, although further clarification of the definition of research-active staff was required. There was less support for the non-portability of research outputs, particularly since the rules for REF2021 were published in 2019 (REF, 2019a), when universities were half-way through the REF cycle. Following the consultation period, specific changes were made to staffing and output submission rules for REF2021 (see Table 2). Further minor adjustments were also made during the assessment process, to reflect the impact of COVID-19 on submitting institutions (see REF, 2020).

One of the key concerns about REF is its cost and burden on the assessment panels.⁸ The reduction in the number of required outputs (from

⁷In fact, the earliest research assessments in 1986 and 1989 were based on outputs generated by the institution.

⁸Technopolis (2015) estimated the total costs of REF2014 as being £246 million, with the costs associated with panellists of £19 million.

4 to 2.5 per FTE) was intended to keep workloads for the assessment panels approximately the same as in REF2014, given expectations that the inclusivity principle and natural sector growth would result in an increase in staff submitted. Additional arrangements were also put in place to support the submission and assessment of interdisciplinary research, overseen by the Interdisciplinary Advisory Panel (IDAP). Furthermore, in response to the documented gender selectivity biases identified in HEFCE (2015), an increased focus was placed on equality and diversity issues, overseen by the Equality and Diversity Advisory Panel (EDAP). The effectiveness of these arrangements are reported in EDAP (2022) and IDAP (2022).

Assessment processes and working methods

REF2021 proceeded in three main phases: (i) the criteria-setting phase (2017–2019) included the sector consultation process following the Stern Report and the final development of the assessment rules; (ii) the submission phase (2020–2021) provided time for institutions to finalize their submissions and for arrangements for the assessment phase to be put in place; and (iii) the assessment phase (2021–2022) involved the sub-panels assessing the submissions for their subject area.

Sub-panel membership and expertise

The selection of panel members commenced with the appointment of the main and sub-panel chairs by Research England, drawing upon nominations from learned associations and other bodies. The 34 subject-based units of assessment were grouped under four main panels. B&M falls under Main Panel C, Social Sciences. These main panels oversaw the work of the sub-panels, provided guidance, checked the assessment criteria were correctly applied and signed-off the results. The appointment of a selection of sub-panel members for the criteria-setting phase followed, which allowed for refinements to the ‘Panel Criteria and Working Methods’ set out by the main panel. Further sub-panel members were subsequently appointed to support the assessment phase, ensuring sufficient, relevant expertise was available across the B&M subject areas and that the panel was di-

verse in terms of individual characteristics, institution types and geography. The final sub-panel for B&M (unit of analysis 17) consisted of 38 ‘full’ sub-panel members, 12 impact assessors and 2 outputs assessors. Eight members had been involved in REF2014, ensuring some continuity of experience. Two sub-panel members were appointed as interdisciplinary advisers, and support was also provided by Main Panel C through an international adviser, an impact adviser and a user adviser. A sub-panel executive comprising of the Chair and two Deputy Chairs led the process, supported by three members of the Secretariat, including one sub-panel adviser.⁹

Sub-panel working methods

All sub-panel members and assessors undertook training on equalities and diversity and unconscious bias. They also participated in workshops to guide them through the assessment rules, processes and IT requirements. Full sub-panel members were involved in assessing outputs, impact case studies and environment statements, while output and impact assessors focused on their respective parts of the assessment. Calibration processes, including training workshops and on-going quantitative and qualitative checks, were embedded throughout the assessment process for all three elements of the assessment. Outputs were assessed by panel members with relevant subject expertise. Impact and environment statements were assessed in three clusters, with approximately 36 institutions allocated to each cluster. Through careful allocation of sub-panel members to these clusters and by instigating regular cross-cluster calibration checks, the sub-panel was able to work efficiently, while ensuring no conflicts of interest arose. During the course of the assessment, sub-panel members were able to raise audit queries on any aspect of the submissions with the secretariat, which was either resolved or raised with the REF administration, on occasions seeking clarification from the submitting institution.¹⁰ The COVID-19

⁹Details of sub-panel membership for SP17 are available at <https://www.ref.ac.uk/panels/panel-membership/>

¹⁰Audit queries were raised with the REF Director when there was insufficient evidence of a substantive connection being established by the supporting statement for submitted staff on minimum fractional contracts (0.20 to 0.29 FTE). A small number of submitted staff were

pandemic had a significant impact on the working practices of the sub-panel. Most meetings involving plenary sessions took place virtually rather than face-to-face, with some hybrid arrangements becoming possible towards the end of the assessment period.

Assessing the submission elements: Allocation and quality assurance

Outputs

The sub-panel received 16,038 outputs, of which 43 were double-weighted. Taking out duplicate outputs (i.e. where the same output was submitted more than once), this provided 13,043 unique outputs. Outputs were matched to the expertise of sub-panel members through an iterative process involving a combination of keywords in titles, journal titles and scrutiny of the outputs themselves. There was no allocation of blocks of outputs by institutions to the same individuals, thus reducing risk of unconscious bias. Members followed the REF guidelines that they were not to use any journal lists or metrics to judge output quality.¹¹ Outputs were assessed normally by one person on a five-point scale (0, 1, 2, 3, 4), with 4 equating to quality that is world-leading in terms of originality, significance and rigour, and 0 equating to the lowest quality standard or not meeting the REF definition of research.¹² Assessors could request the reallocation of outputs that fell outside their expertise, seek guidance from another sub-panel member or request cross-referral of outputs that fell outside the expertise of SP17 to another sub-panel. Where deemed necessary, for example on borderline scores, assessors could discuss the output with another sub-panel expert. The evolving distributions of output scores by panel member were regularly presented to the sub-panel, and through this transparency the final distribution of output scores was similar across panel members and hence across sub-disciplines.

removed from the REF database because they were deemed ineligible.

¹¹Panel Criteria and Working Methods, paragraph 207 (REF, 2019b).

¹²See Guidance on Submissions, Annex A, Table A2 (REF, 2019a) and Panel Criteria and Working Methods, Section 3 (REF, 2019b).

Cross-referrals, joint assessment and interdisciplinary research outputs

The sub-panel received a large number of interdisciplinary and multidisciplinary outputs, with an increase in the number that required cross-referral or joint assessment with other sub-panels. These may have been flagged by the submitting institution or where the expertise needed fell outside that of SP17. The sub-panel cross-referred (or jointly assessed) 2,385 of all outputs submitted to B&M (2,241 of these were unique outputs). As expected, a significant number of cross-referrals fell clearly within the remit of SP16 Economics and Econometrics: 1342 unique outputs were cross-referred to this sub-panel. Given the scale, this particular process was overseen by sub-panel members jointly appointed with B&M (SP17) and Economics and Econometrics (SP16). Other outputs were cross-referred elsewhere: a relatively large number went to SP2 Public Health (120 unique outputs), SP4 Psychology (115), SP14 Geography (101), SP21 Sociology (95), SP23 Education (88), SP18 Law (74) and SP19 Politics (65). The advice and scores received were then reviewed by the B&M sub-panel before final output scores were recorded.

In response to the growth in interdisciplinary research, submitting institutions had been invited to flag 'interdisciplinary' research (IDR) outputs to ensure assessment by suitable sub-panel members. The sub-panel noted that the IDR flag was used extensively by some institutions but not at all by others. Where the same output was submitted by different institutions (duplicates), one may have flagged the output as being IDR while another may not have done so. Outputs flagged as IDR were scrutinized by the two SP17 IDR assessors, who considered whether they should be assessed within SP17, cross-referred or jointly assessed with another sub-panel. Of the 866 outputs flagged as IDR by submitting institutions, 539 were assessed within SP17, with the rest being cross-referred or jointly assessed by other sub-panels.

Impact

The sub-panel received 539 impact cases for assessment. Each case was allocated to a group of three assessors, comprising the primary and secondary assessor of the institution to which the impact cases belonged and an impact assessor. Assessors scored each impact case independently on

a nine-point scale (0, 0.5, 1.0, ..., 3.5, 4.0), before then discussing in their triplet group and agreeing a provisional score.¹³ These scores and their rationale were then subject to scrutiny in wider meetings of the relevant clusters to ensure consistency, before a final score was agreed. Any generic issues regarding the process and alignment of scores were discussed in full sub-panel plenary.

The significance of impact in the assessment is not to be underestimated. The types of impact assessed by the B&M sub-panel were at the organizational, economic and societal levels and spanned private, public and third-sector bodies, as well as public policy design and intervention. Impacts were observed within local, national and international domains, and included a range of collaborations and partnerships. Engagement and impact were clearly an integral and important part of the research strategy and investment of the submitting institutions, and reflected the breadth of research with B&M.

Environment

Each of the 108 institutional environment statements was allocated to the same primary and secondary assessor as with impact. Working independently and then together in pairs, assessors followed the quality criteria to score each of the four elements of the environment based on a nine-point scale (0, 0.5, 1.0, ..., 3.5, 4.0). Having discussed provisional scores and their rationale, these were subject to further review and discussion within each cluster before a final environment profile was developed.¹⁴ The environment statements provided an important context for each submission and included important data and information, for example on governance, policies for ECRs and equality and diversity. Assessors were provided with a university environment statement, to set out the broader context of the submission, as well as the environment statement for the B&M unit. The latter formed the basis for the assessment score. The diversity of the submitting institutions, their various contexts, sizes and presentation of

the information varied considerably. Thus, assessing these statements using the REF criteria was one of the most challenging aspects of the REF. However, through discussions, calibrations exercises and, where necessary, checking of information through audits, consensus was reached amongst the assessors and the sub-panel.

Summary of submissions and quality profiles

Of the 108 submissions, covering 6,638.27 FTE staff (7,008 headcount), there was a widespread distribution of submission sizes, the largest being 186.9 FTEs and the smallest 5.4. The largest 26 submissions accounted for 50% of the total submitted FTEs. 1,031 staff were designated as early career researchers (ECRs), 15% of the headcount. These numbers imply that the average size of the 108 submissions was 61.5 research-active staff. However, the size of UK B&M submissions is skewed, with a moderate number of large institutions and a long tail of smaller institutions. Seventeen institutions submitted nine or more impact case studies; 35 institutions submitted seven or more, accounting for more than half of those assessed (291); while 25 submissions included only two impact case studies. The quality profile for SP17 presented in Table 3 shows that a substantial majority of the submitted material was assessed to be of 4* or 3* quality.

Table 4 provides a comparison with data from REF2014. Care needs to be exercised when making comparisons and drawing conclusions because the rule changes between REF2014 and REF2021 will have affected both the scale and outcomes of the exercise. For example, output profiles are likely to have been affected by the rule changes around non-selectivity and the reduction in average number of outputs submitted per FTE.

Submissions to REF2021 included 10 new entrants, with three institutions of the 101 that submitted in 2014 choosing not to make a submission in 2021. The number of staff submitted in 2021 was more than a 100% increase from the 3,300 staff submitted in 2014, reflecting both the growth in the discipline and the REF2021 rule requirement of no selectivity. The 16,038 outputs compared with 12,204 in 2014: a rise of 32%. The scale of this rise in volume of submitted material was not unexpected and nor was it the intention of the change in

¹³See Guidance on Submissions, Annex A, Table A3 for the criteria and definitions of quality levels for assessing impact (REF, 2019a).

¹⁴See Guidance on Submissions, Annex A, Table A4 for the criteria and definitions of the quality levels for assessing environment (REF, 2019a).

Table 3. Quality profile for business and management studies SP17: FTE weighted

Profile type	Weight in overall profile (%)	4* (%)	3* (%)	2* (%)	1* (%)	Unclassified (%)
Output	60	30.4	45.9	21.3	2.1	0.3
Impact	25	41.7	42.8	13.4	2.1	0
Environment	15	42.4	40.7	15.0	1.9	0
Overall	100	35	44	19	2	0

The table shows the distribution of quality profiles overall and for outputs, impact and environment scores separately for SP17 in REF2021 weighted by FTE numbers.

Table 4. Quality profiles in business and management studies: REF2021 and REF2014

Profile type	Year	Number of FTE staff	Number of assessments	4* (%)	3* (%)	2* (%)	1* (%)	Unclassified (%)
Overall	2021	6,638		35	44	19	2	0
	2014	3,300		26	43	26	4	1
Output	2021		16,040	30.4	45.9	21.3	2.1	0.3
	2014		12,204	20.5	42.8	30.1	5.8	0.8
Impact	2021		539	41.7	42.8	13.4	2.1	0
	2014		432	37.7	42.5	17.0	2.2	0.6
Environment	2021		108	42.4	40.7	15.0	1.9	0
	2014		101	36.8	39.7	21.0	2.4	0.1

The table shows the distribution of quality profiles for outputs, impact and environment for SP17 in REF2021 compared with the previous REF2014. Care must be taken in comparing the data because of changes in the rules for submission. See text for discussion.

required outputs to 2.5 per FTE.¹⁵ Rather, this non-selectivity rule change had been intended to preserve the volume of work being assessed. The increase in staff, outputs and impact cases submitted to B&M between 2014 and 2021 was high compared with Main Panel C (growth in FTE staff submitted excluding B&M was 51.6%) and very high compared with the other three main panels (growth in FTE staff submitted was 39.9%). This demonstrates the significant contribution of B&M research to the social sciences and the university sector in general.¹⁶

Although the changes in submission rules mean comparisons between REF2014 and REF2021 in Table 4 should be treated with caution, this improvement in quality suggests that UK research in B&M has not been diminished by the increase in volume of staff and research submitted. Overall, in 2021, 35% of the material submitted was judged to be world-leading and 44% internationally excellent, representing a substantial improvement on

REF2014, which highlights the healthy state of UK business and management research.

Analysis of output scores

In analysing output scores, the sub-panel had a series of issues to address. The REF rules allowed for the same output to be submitted more than once. Duplicate outputs could occur for two reasons: first, because the output had multiple co-authors and each co-author at a different institution may have submitted the same output; and second, because under the no-portability transition arrangements, authors may have moved institutions, resulting in multiple institutions submitting the same output on behalf of the same author. The REF rules required that all duplicate outputs (those with the same DOIs) were awarded the same grade.¹⁷ Of the 16,038 outputs, 2,161 were submitted more than once, with one output submitted nine times! The output profiles

¹⁵Although the submission intentions in September 2019 had alerted REF to the likely increase in B&M outputs.

¹⁶Indeed, SP17 had the largest number of impact cases and institutions of all the sub-panels across REF2021.

¹⁷There was no requirement for the same output submitted to different sub-panels to be awarded the same grade, on the basis that the quality of a specific output might be judged differently by different sub-panels within their subject area.

Table 5. Output quality profile of unique outputs in business and management

Sample	Number	Mean GPA	St. dev. GPA	4* (%)	3* (%)	2* (%)	1* (%)	Unclassified (%)
Section A: Full sample of unique outputs								
All unique outputs	13,403	3.00	0.787	28.5	46.0	22.9	2.5	0.1
Section B: Cross-referred/joint outputs								
SP17 outputs	11,162	3.02	0.797	29.9	44.6	22.9	2.5	0.1
CR outputs	2,241	2.94	0.734	21.8	53.0	22.9	2.3	0.0
t-Test difference in means (p-value)	4.068** (0.000)							
Section C: Interdisciplinary outputs								
Non-ID outputs	13,242	3.00	0.788	28.6	45.9	23.0	2.5	0.1
ID outputs	161	2.91	0.705	17.4	59.0	20.5	3.1	0.0
t-Test difference in means (p-value)	1.577 (0.115)							
Section D: Co-authored outputs								
Single authored	1,285	2.91	0.812	25.5	44.1	26.9	3.58	0.0
Multiple authors	12,118	3.01	0.784	28.8	46.2	22.5	2.4	0.1
t-Test difference in means (p-value)	-4.30** (0.000)							

The table shows the quality profile of panel scores awarded to outputs, the grade point average (GPA) and variability in these scores. Section A reports the profile for all unique outputs; Section B compares outputs assessed within SP17 with those outputs that were cross-referred; Section C compares outputs that were classified as interdisciplinary with non-interdisciplinary outputs; Section D compares outputs with single and multiple co-authors. Using these mean GPAs, comparisons are made via a simple t-test of differences in means. ** denotes significance at the 99% confidence level. Numbers in this table include only unique outputs.

reported in Table 4 include the duplicate outputs. Once these duplicate outputs are removed, there were 13,403 unique outputs. Table 5 (Section A) reports the quality profile of these unique outputs and also the grade point average (GPA) associated with the quality distribution.

As can be seen, the percentage of unique 4* outputs is slightly lower than in Table 3, reflecting that outputs that had been submitted multiple times tended to be awarded slightly higher grades. A t-test confirmed the difference in mean GPA between the 11,242 outputs submitted only once and the 2,161 outputs submitted multiple times.

SP17 retained responsibility for the final scores for all outputs originally allocated to it, taking into consideration the detailed advice received from the respective cross-referred sub-panel. Table 5 (Section B) reports the quality profile of the unique outputs assessed within SP17, compared with those outputs that were either cross-referred (2,216) or jointly assessed (25). The analysis shows that cross-referred outputs had a smaller percentage of 4* grades and a slightly lower GPA: results that are statistically significant. The sub-panel examined these differences and following various

checks to ensure that cross-referred outputs had been fairly assessed, were reassured that this was the case.

Turning to IDR outputs, Table 5 (Section C) compares the quality profiles of 161 unique IDR outputs and the remaining 13,242 unique outputs. Although IDR outputs received fewer 4* grades, they also received a higher percentage of 3* grades and fewer 2* grades than non-IDR outputs. The point estimates of the GPAs suggest that IDR outputs had a lower average grade but the difference was not statistically significant at conventional confidence limits. Hence, overall, IDR outputs received similar quality grades to non-IDR outputs.

A range of output types was submitted, with the overwhelming majority being journal articles. The majority of unique outputs were written by multiple authors, most commonly having between two and four co-authors (81% of all unique outputs), with 1,120 outputs having more than four co-authors. Many of these outputs were international or emerged from international collaborations. There appears to be some value in involving co-authors in an output in terms of enhanced

Table 6. GPA score in business and management studies by unique output type

Output type	Number	Mean GPA	St. dev. GPA
Book (authored/edited) (38 double-weighted)	191	3.00	0.921
Chapter in book	110	1.99	0.829
Journal article	12,974	3.02	0.774
Conference contribution	42	1.88	0.861
Research report for external body (2 double-weighted)	25	2.24	0.970
Working paper & other	61	2.70	0.882

The table shows the average REF panel score (GPA) awarded to different types of output (books, book chapters, journal articles, conference papers, research reports and working papers) and the variability of these scores. Numbers in this table include only unique outputs.

quality. Table 5 (Section D) compares the GPA of the 1,285 single-authored outputs with the 12,118 outputs with multiple authors (where these details had been provided by the submitting institutions). There is a statistically significant difference in the GPA of +0.10 for multiple-authored outputs.

One of the effects on business and management of successive research assessments has been the move towards peer-reviewed journal articles, even though the REF rules indicate that all output, irrespective of type, is assessed using the criteria of originality, significance and rigour (REF, 2019a). Table 6 lists the types of outputs submitted to SP17, together with their corresponding GPA score.

Book chapters, external reports and conference contributions continue to be minority forms of output submitted for assessment (Table 6). The GPA for authored books almost matched the GPA score for journal articles, with 20% of these outputs being double-weighted. However, all other forms of output received lower grades, with conference contributions and book chapters receiving the lowest average scores. Thus, whilst the emphasis on submitting peer-reviewed journal papers may be an appropriate, though conservative, submission strategy, the low number of books submitted may be worthy of reflection by those in universities making decisions about what types of output to submit.

Inclusivity

The sub-panel was supplied with data on the distribution of outputs across staff and by ECR and former staff, as summarized in Table 7. One of the key features of REF2021 was the emphasis on inclusivity and we have investigated this ambition using these data.

The total headcount of submitted staff to B&M was 6,952, including 982 ECRs and 864 former staff for whom outputs were submitted. When the 1,546 outputs submitted by the 864 former staff are added to the 14,557 outputs submitted by the 6,952 head count, this yields the total of 16,103 outputs originally submitted.¹⁸ Table 7 also shows that 3,373 staff (48.5% of the total headcount) submitted a single output but these accounted for only 23.2% of all submitted outputs. In contrast, 638 staff (9.2% of headcount) submitted five outputs representing 21.9% of outputs. For ECRs, 693 (70.6%) individuals submitted a single output, with just 21 of this group (2.1%) submitting five outputs. This analysis demonstrates that research outputs are concentrated amongst a relatively small number of staff, with 57.3% of outputs submitted by 31% of staff. These results may be interpreted in different ways. First, the fact that a significantly large number of staff submitted one output may be regarded as the REF achieving its ambition of inclusivity. The large number of ECRs submitted to the REF may be regarded as a measure of success. On the other hand, the importance of a small proportion of publishing 'stars', contributing to a large share of outputs submitted, may suggest that inclusion remains an ambition to pursue and yet be realized.

REF assessment of quality and journal rankings

All academics working in business and management schools will be aware of the lists of academic journals ranked by quality (Harzing, 2022). The

¹⁸These numbers refer to submission data on 31 March 2021 and are different from subsequently reported numbers due to exclusions for various reasons; for example, a reserve output for double-weighted requests.

Table 7. Distribution of number of attributed outputs in business and management SP17

Number of outputs attributed	Headcount	Percent of headcount	Number of outputs by headcount	Percent of outputs	Former staff headcount	Number of outputs by former staff	ECR headcount	Number of outputs by ECR
1	3,373	48.5%	3,373	23.2%	485	485	693	693
2	1,422	20.5%	2,844	19.5%	197	394	168	336
3	926	13.3%	2,778	19.1%	92	276	65	195
4	593	8.5%	2,372	16.3%	59	236	35	140
5	638	9.2%	3,190	21.9%	31	155	21	105
Total	6,952	100.0%	14,557	100.0%	864	1,546	982	1,469

The table shows the number of outputs submitted attributed by headcount of staff (ECRs and former staff members) across all SP17 outputs and submitted staff members. Numbers in this table were correct at date of submission on 31 March 2021.

use and value of these ranking schemes in business schools is analysed by Walker *et al.* (2019), together with a critique by Mingers and Willmott (2013). Others have expressed concern about the origin and use of ‘elite’ journal lists, predominantly from the Western world, and their adverse implications for sub-disciplines within business and management, scholars worldwide and emerging academic communities (e.g. IFSAM, 2021). As was made clear in the REF Guidelines paragraphs 2 and 17, the REF is a process of expert review, and the Panel Criteria and Working Methods document (REF, 2019b), paragraph 207, states that ‘No sub-panel will use journal impact factors or any hierarchy of journals in their assessment of outputs’. In the run-up to the submission date, the B&M sub-panel chair and sub-panel members emphasized this position and reassured the community that no form of metric would be used in the assessment process. Thus, the REF process relies on expert assessment through applying the criteria of originality, significance and rigour, which may be different from the review and criteria utilized by journals.

Pidd and Broadbent (2015, table 3) undertook a comparison of a sample of 1,000 outputs graded in REF2014, with journals ranked by the Chartered Associate of Business Schools (CABS) Academic Journal Guide (AJG), finding that about half of the REF output scores matched with the CABS list. We have developed a similar analysis for REF2021 using the REF2021 unique output data. We used details and ratings of all of the journals listed in the AJG 2015, as this was close to the start of the REF period and may have influenced the choice of publication outlet by submitting institutions, research managers and authors.¹⁹ Of the 12,974 unique journal outputs submitted to REF2021, we were able to obtain a match between 11,922 outputs and their AJG rating (some submitted outputs were in journals not in the AJG).

A simple correlation coefficient between these two sets of scores is 0.466. An interpretation of this correlation coefficient is that the coefficient of determination is 0.221 and 22.1% of the variation in panel scores is explained by the AJG ranking. The contingency table shown in Table 8 details the AJG ratings and the awarded REF panel scores.

¹⁹For details of the CABS Academic Journal Guide, see CABS (2021).

Table 8. Contingency table of CABS AJG ratings and realized REF panel scores

Panel score	ratings_2015				Total
	1	2	3	4	
0	0	0	2	0	2
	0	0	100	0	100
	0	0	0.04	0	0.02
1	25	58	69	23	175
	14.29	33.14	39.43	13.14	100
	19.84	7.56	1.26	0.41	1.47
2	70	402	1,558	474	2,504
	2.8	16.05	62.22	18.93	100
	55.56	52.41	28.49	8.52	21
3	28	268	3,110	2,235	5,641
	0.5	4.75	55.13	39.62	100
	22.22	34.94	56.88	40.19	47.32
4	3	39	729	2,829	3,600
	0.08	1.08	20.25	78.58	100
	2.38	5.08	13.33	50.87	30.2
Total	126	767	5,468	5,561	11,922
	1.06	6.43	45.86	46.64	100
	100	100	100	100	100

The table shows the relationship between the outputs classified by AJG ratings and the awarded REF score. In each block, the first number denotes frequency, the second number denotes row percentage and the third number denotes column percentage. Numbers in this table include only unique outputs which match with journals in the AJG list. The elements highlighted in bold in the shaded boxes on the leading diagonal show the number of outputs (and related percentages) where the AJG ratings and the REF scores are the same; the off-diagonal elements show the numbers and percentages where there are differences.

In each block of Table 8 there are three sets of numbers, the first is the frequency of outputs in that particular block. These frequency numbers in the Total column show that the sub-panel graded 3,600 outputs out of 11,922 as 4* and at the lower end, 175 as 1*. Table 8 also shows that 2,829 outputs out of the 3,600 that were scored 4* by the REF panel were also in journals graded 4* by the AJG ratings. The second number in each block reports the percentage of outputs in a REF scoring category that were in an AJG rating category. For example, 78.58% of outputs scored 4* in the REF were in journals rated 4* by the AJG. The horizontal sum of the second set of numbers is 100% in the final column. The third number in each block shows the percentage of outputs in an AJG rating category that are in a REF scoring category.

Overall, we find that 50.87% of outputs in 4*-rated journals in the AJG also achieved a REF score of 4*. The interpretation of this number is that 0.51 is the probability that an output obtains a REF score of 4*, conditional on this output being in a journal designated by the AJG as being 4*. The elements in bold down the leading diagonal (shaded boxes) of Table 8 report the AJG ratings

and realized REF scores being the same; the off-diagonal elements show those cases where the REF scores are different from the AJG ratings. To focus on the high end of the quality spectrum, we further identified 2,843 of the 11,922 journal outputs that were in the Financial Times top 50 journals: sometimes considered the 'elite' publication outlets in business and management. Of these 2,843 outputs, 61.3% received a 4* panel score, 32.0% received a 3* score and 6.5% received a 2* score.

The above analyses show that the relationship between journal rankings and the assessed scores for individual outputs using REF criteria is unstable. It adds further weight to the argument that peer review is critical in the REF process, particularly when judging against the quality criteria of originality, significance and rigour (Frey and Rost, 2010). The results also imply that in making their REF decisions on what to submit, research managers should not rely on journal rankings alone as a means of output selection.²⁰

²⁰Note that journal guides and their metrics may not be intended for use in making decisions about what to submit to the REF. Rather, they are often used by researchers in making decisions on where to publish and what to include.

Analysis of impact cases and environment statements

The profiles in Tables 1 and 2 show a slight improvement in the quality of impact cases assessed, and in the quality of the environment statements. Overall, 41.7% of impact cases were assessed as 4* (37.7% in REF2014) and 42.4% of environment statements as 4* (36.8% in REF2014). Kellard and Sliwa (2016) analysed B&M impact cases from REF2014, focusing on the organizational contexts generating impact, and concluded that impactful research is linked to the researcher's long-term association with the submitting institution. Thorpe *et al.* (2018) examined B&M environment statements from REF2014 in terms of language-related characteristics between high-ranked and low-ranked submissions and note the difficulties in assessing textual claims.

As explained above, impact cases and environment statements in REF2021 were assessed by multiple panel members who provided provisional scores that were then discussed in clusters before the final scores were agreed. Here we extend the REF analysis by examining whether the impact scores and environment scores awarded by SP17 are related to various HEI characteristics. Our analysis explores the outcomes of these assessments with regard to specific institutional characteristics as independent variables that were not an explicit part of the REF quality assessment criteria. As stated in the REF Working Methods: 'The main panels have no preformed view of the ideal size or organisational structure for a research environment, or of the ideal context or approach for enabling impact, and will judge each submission on its merits, contextualised appropriately to the nature of institution' (REF, 2019b, paragraph 336). The purpose of our analysis is to examine the relationship between these characteristics and the REF GPA outcomes.

Table 9 reports the results of ordinary least squares (OLS) regressions for the variables that potentially explain impact GPA and environment GPA for the 96 institutions in REF2021 that were also submitted in REF2014. The dependent variables for each institution for impact and environment were obtained by taking the profile awarded and converting it into a GPA.

The first column in Table 9 reports the effect of institution characteristics on the impact GPA, and the first characteristic is the size of the submission

Table 9. Institution characteristics and impact and environment scores

	Dependent variable	
	Impact_GPA_2021	Env_GPA_2021
Size_FTE	0.003** (0.001)	0.017** (0.004)
Size_squared		-0.00006** (0.00002)
% change in FTE	-0.114** (0.033)	-0.111** (0.04)
PhD completions (per_FTE)		-0.061* (0.027)
Res_income (£ per FTE)		0.46 × 10 ⁻⁵ (0.4 × 10 ⁻⁵)
GPA_2014	0.223** (0.070)	0.344** (0.087)
Intercept	2.376** (0.198)	1.445** (0.192)
Observations	96	96
F statistic	15.28	47.23
R-squared	0.33	0.76

The table reports the results of the OLS regressions, with results from Tobit regressions being very similar. Standard errors are reported in brackets. ** and * denote significance at the 99% and 95% confidence intervals, respectively. The reported regressions only include those variables that were significant. Insignificant variables not reported include indicator variables for panellists (which were jointly insignificant), indicators for clusters, proportions of FTE staff, ECR staff, eligible staff, former staff and research income in REF2014.

based on the number of FTE staff. The coefficient on this variable is positive and significant, showing that for larger submissions, although the number of impact cases submitted increased, the average quality of these cases also increased. Of course, the size of a submission masks a myriad of factors, including organizational structures, strategies, resources, levels of autonomy and reputation. Therefore, care needs to be taken in interpreting the relationship between size and GPA performance.

What of the effects of the dynamics of change in an institution's submission? Our analysis shows that the percentage change in the number of FTE staff between REF2014 and REF2021 has a significantly negative value, suggesting that growth in staff submitted negatively affects impact performance. Finally, we find that the impact GPA awarded in REF2014 and REF2021 has a significantly positive coefficient. This suggests some persistence in the quality of impact cases, with institutions that scored high or low with impact

in REF2014 continuing to score high or low in REF2021.

The second column of Table 9 reports on the relationship between certain characteristics of submitting institutions and the environment GPA scores. The size of the submission, in terms of FTE staff, had a significantly positive effect on the panel score awarded, suggesting the environments of large submissions are assessed as stronger in terms of their vitality and sustainability. However, the coefficient on the quadratic size term is negative and significant, indicating that the benefits of size decline and tail off in very large submissions. In terms of changes between 2014 and 2021, the third row shows that the effect of an increase in the percentage of FTE staff size of the submission on the quality profile from 2014 to 2021 is negative and significant. One reason for this relationship may be that a growth in staff numbers harms the research environment. Another may be that those institutions that were highly selective of who to submit in 2014 and had to increase their submission numbers to meet the REF2021 no-selectivity principle were relatively weaker than those that had previously submitted a larger proportion of their staff in 2014. Surprisingly, the number of doctoral completions per FTE has a negative and significant coefficient. This may reflect the strategies followed by some institutions to limit their PGR student intake, thus reducing their completions per FTE. On the other hand, over-recruiting of PGR numbers may lead to untenable high numbers of completions per FTE and result in a downward pressure on the research environment assessment score. Research income per FTE has a positive, although not significant, effect on the environment score. Finally, there is a positive and significant relationship with the previous environment scores of REF2014, suggesting some persistence in the quality of research environment.

We also examined raw correlations between an institution's output, impact and environment GPAs. The correlation across institutions between outputs and environment was 0.81; the correlation (impact, environment) was 0.66; and the correlation (outputs, impact) was 0.64. If an argument is made for causality from environment through to outputs, then the square of the correlation coefficient (0.65) can be interpreted as the coefficient of determination in a regression and would imply that institutional environment GPAs explain 65% of the variability across institutions in out-

put GPAs. Similarly, environment GPAs explains 44% of the variability in impact GPAs. It is more difficult to argue for a causal relationship between outputs and impact because outputs were assessed over 2014–2020 whilst the underlying research output for impact cases could go back to 2000. These high correlations echo the analysis by Pinar and Horne (2022) of REF2014 data across all panels on the value of the separate elements of the REF. However, we would be cautious about extrapolating these correlations to the idea of removing an element or elements of the REF, as this would most likely lead to a change in behaviour and a narrowing of universities' definition of research and a re-focusing of their priorities.

Conclusions and implications

In this paper we have outlined the processes and outcomes of the latest UK research evaluation, REF2021. Along with other sub-panels in REF2021, we have produced a detailed report of the processes and outcomes of the REF which we do not seek to replicate here (see REF, 2022b). In outlining the evaluation processes and providing additional analyses of the results, we have aimed to stimulate discussion and debate within the B&M community and higher education as a whole. We draw the following key messages on assessing research quality, in relation to the REF in general as well as specifically for B&M studies.

First, we have provided evidence that journal lists and citation indices do not correlate with the peer assessment of research quality.²¹ We add to those voices arguing that the automation of research quality assessment, using metrics, is inappropriate and concur with others that the REF's use of expert review is an appropriate, although not perfect, means of assessing research quality (Geuna and Piolatto, 2016). In making this comment, it is not our intention to criticize ranking schemes or guides. Nor do we advocate, either for or against, their use by academics and their institutions in informing where to submit their work. Rather, we point out that these lists are not designed to fulfil the requirements of the REF, which aims to assess output quality on the basis of originality, significance and rigour. Hence, whilst we

²¹ This evidence is consistent with the Declaration on Research Assessment (DORA).

concur with the view that ‘... there is no true state of nature for any of these papers’ (Pidd and Broadbent, 2015, p. 574), we reinforce the position that reliance on journal lists or metrics for assessing quality in B&M research outputs would be at best misleading and at worst dangerous (Mingers and Willmott, 2013).

Second, the change in REF2021 rules around non-selectivity of staff and greater inclusivity had several effects on B&M submissions and quality outcomes. Greater staff inclusivity was evident in the increased size of institutions’ submissions, in the substantial increase in the research elements of B&M as a whole and from the numbers of included ECRs. Furthermore, the change in output requirements from 4 to 2.5 per FTE allowed institutions greater flexibility to shape their submissions to reflect factors such as career stage. Although this move will help to reduce the stigma associated with whether an academic’s work has been submitted or not to the REF, we hope that it will not be replaced by a new stigma linked to the number of outputs an academic has returned. Even though the REF aims to eschew exclusivity and the establishment of an elite set of institutions, our analysis of specific institutional characteristics and REF outcomes suggests some unintended consequences. We have noted the concentration of submitted outputs from a relatively small number of academics which, in part, is a consequence of the portability transition arrangements. We have also reported that the overwhelming output types submitted were journal articles, and there is a concern that other types of output – such as books and research monographs – are squeezed out of the set of admissible output types. Submissions to the B&M sub-panel also ranged considerably in size, with very small submissions sometimes struggling to demonstrate vitality and sustainability in their research environments or, having limited staff resources, found it difficult to satisfy the minimum requirement of two ICS. For institutions that previously had been very selective in terms of the staff included, this effect could be compounded by requiring a larger number of ICS compared with REF2014. Conversely, our analysis suggests that the environment criteria for assessment appear to benefit larger submissions or, at least, that there could be a minimum efficient and optimum size of institution. Hence, inclusivity is a desirable and laudable ambition but it brings with it some challenges for particular institutions.

Third, there are implications of the intellectual costs and benefits of the exercise, particularly in such a large submission as B&M. We have noted throughout the considerable administrative burden of the whole process in terms of the preparation of submissions and submitting institutions’ infrastructures, the central REF administrative structures and the assessment procedures. Such costs have consequences for the future resource-intensiveness of research evaluation exercises. We have commented on the size of the B&M sub-panel in relation to other sub-panels within the social sciences and the mitigatory actions taken to manage this scale when assessing research environment and impact. This raises the question of whether the REF processes are appropriate for such a wide variation in panel sizes. However, whilst we recognize the multi-faceted costs associated with the REF and accept the need for reform, calls for removing specific elements (environment and/or impact) of submissions should be resisted as this may lead to deleterious effects on those activities not subject to assessment and ultimately diminish the impact of B&M research on the environment, society and the economy.

Fourth, the research councils emphasize the importance of interdisciplinary research and in this REF, there was an attempt to ensure that IDR work was not disadvantaged in its assessment. However, difficulties remain in the assessment of IDR research, and the best process for assessing and hence incentivizing academics to undertake IDR research is still unclear. IDR advisers cannot be expected to be experts across all 34 units of assessment. Much research and impact in B&M is inherently interdisciplinary in its research methods and subject material, particularly when impact cases are involved (MacIntosh *et al.*, 2021). The growing emphasis placed by many business and management schools on research that addresses wider societal and economic problems is one factor driving this trend, as is the tendency to work with researchers from other disciplines and B&M sub-disciplines. Flagging IDR research outputs was, therefore, problematic for institutions because of the already interdisciplinary nature of the work within business and management. This was further confused because duplicate outputs were often flagged differently by different institutions. If submitting institutions cannot agree whether their work is interdisciplinary, it makes it challenging for the sub-panel to identify appropriate assessors.

Thus, IDR is taking place within B&M studies, but the concept requires more consideration and clarification when designing assessment frameworks. Of course, these observations from B&M should not detract from the excellent IDR outputs the sub-panel received, but the points raised require verification within other units of assessment and remain open for debate.

Finally, this paper has demonstrated the intellectual vibrancy of research in B&M studies, although we have been unable to do justice to the range of subject areas addressed.²² The internationalization of UK B&M studies research is also evident in all three elements of the assessments, including in the extent of international networking and co-publishing, and the location and reach of impact. The databases produced by the REF are publicly available and should be regarded as a resource for researchers and users of research, illuminating the quality and impact of research. These show clearly the variety of subjects, the relevance, international reach and high quality of the research undertaken by B&M scholars.²³ The high-quality environments in which B&M scholars pursue their research agendas, the breadth and strength of the impact cases and the proportion of internationally excellent and world-leading outputs demonstrate a thriving field of endeavour.

References

- Basso, A. and G. di Tollo (2022). 'Prediction of UK research excellence framework assessment by the departmental *h*-index', *European Journal of Operational Research*, **296**, pp. 1036–1049.
- BEIS (2016). *Building on Success and Learning from Experience: An Independent Review of the Research Excellence Framework* (Stern Report). London: BEIS.
- Bessant, J., S. Birley, C. Cooper, S. Dawson, J. Gennard, M. Gardiner, et al. (2003). 'The state of the field in UK management research: reflections of the Research Assessment Exercise (RAE) panel', *British Journal of Management*, **14**, pp. 51–68.
- BIS (2015). *Fulfilling our Potential: Teaching Excellence, Social Mobility and Student Choice*. London: BIS.
- Bleiklie, I. and S. Michelsen (2015). *Political-administrative regimes and university governance*. Available at: <https://ecpr.eu/Filestore/PaperProposal/48e6c9c2-7460-43a9-aaf7-557bdb35e4cf.pdf>. last accessed 8-10-2015
- CABS (2021). *Academic Journal Guide*. Available at: <https://chartereddabs.org/academic-journal-guide-2021/>
- Cooper, C. and D. Otley (1998). 'The 1996 research assessment exercise for business and management', *British Journal of Management*, **9**, pp. 73–89.
- De Boer, H., B. Jongbloed, P. Bennenworth, L. Cremonini, R. Kolster, A. Kottmann, et al. (2015). *Performance-Based Funding and Performance Agreements in Fourteen Higher Education Systems*. Report for the Ministry of Education, Culture and Science. The Hague: Ministry of Education, Culture and Science.
- Dougherty, K. J., S. M. Jones, H. Lahr, R. S. Natow, L. Pheatt and V. Reddy (2016). *Performance Funding for Higher Education*. Baltimore, MD: Johns Hopkins University Press.
- EDAP (2022). *Equality and Diversity Advisory Panel Final Report*. Available at: <https://www.ref.ac.uk/publications-and-reports/>
- Franco-Santos, M. and D. Otley (2018). 'Reviewing and theorizing the unintended consequences of performance management systems', *International Journal of Management Reviews*, **20**, pp. 696–730.
- FRAP (2022). *Future Research Assessment Programme*. Available at: <https://www.ukri.org/about-us/research-england/research-excellence/future-research-assessment-programme-frap/#contents-list>
- Frey, B. S. and K. Rost (2010). 'Do rankings reflect research quality?', *Journal of Applied Economics*, **13**, pp. 1–38.
- Geuna, A. and B. R. Martin (2003). 'University research evaluation and funding: an international comparison', *Minerva*, **41**, pp. 277–304.
- Geuna, A. and M. Piolatto (2016). 'Research assessment in the UK and Italy: costly and difficult, but probably worth it (at least for a while)', *Research Policy*, **45**, pp. 260–271.
- Harzing, A.-L. (2022). *Journal Quality List, 69th Edition*. Available at: <https://harzing.com/resources/journal-quality-list>
- HEFCE (2015). 'Selection of staff for inclusion in the REF2014', Issues Paper, August 2015/17.
- Hicks, D. (2012). 'Performance-based university research funding systems', *Research Policy*, **41**, pp. 251–261.
- Hicks, D., P. Wouters, L. Waltman, S. De Rijcke and I. Rafols (2015). 'Bibliometrics: the Leiden Manifesto for research metrics', *Nature*, **520**, pp. 429–431.
- Hood, C. (1995). 'The "new public management" in the 1980s: variations on a theme', *Accounting, Organizations and Society*, **20**, pp. 93–109.
- IDAP (2022). *Interdisciplinary Research Advisory Panel: Final Report*. Available at: <https://www.ref.ac.uk/publications-and-reports/>
- IFSAM (2021). *IFSAM General Assembly's Position Statement on Management Research*. Available at: <https://www.ifsam.org/blog/2021/06/04/ifsam-general-assemblys-position-statement-on-management-research/>
- Jonkers, K. and T. Zacharewicz (2016). *Research Performance-Based Funding Systems: A Comparative Assessment*. Luxembourg: Publications Office of the European Union.
- Kellard, N. and M. Sliwa (2016). 'Business and management impact assessment in Research Excellence Framework 2014: analysis and reflection', *British Journal of Management*, **27**, pp. 693–711.
- MacIntosh, R., K. Mason, N. Beech and J. M. Bartunek (2021). *Delivering Impact in Management Research: When Does it Really Happen?* Oxford: Routledge.

²²For a detailed review of subject areas within B&M, see the full sub-panel report (REF, 2022b).

²³For example, an analysis of the B&M impact cases by the Rotterdam School of Management demonstrates a strong alignment with specific UN Sustainability Goals (<https://rsmmetrics.nl/sustainable-development-goals/ref-impact-cases/>).

- Mingers, J. and H. Willmott (2013). 'Taylorizing business school research: on the "one best way" performative effects of journal ranking lists', *Human Relations*, **66**, pp. 1051–1073.
- Pidd, M. and J. Broadbent (2015). 'Business and management studies in the 2014 Research Excellence Framework', *British Journal of Management*, **26**, pp. 569–581.
- Pinar, M. and T. J. Horne (2022). 'Assessing research excellence: evaluating the Research Excellence Framework', *Research Evaluation*, **31**, pp. 173–187.
- REF (2019a). *Guidance on Submissions*. Available at: <https://www.ref.ac.uk/guidance-and-criteria-on-submissions/guidance/>
- REF (2019b). *Panel Criteria and Working Methods*. Available at: <https://ref.ac.uk/publications-and-reports/panel-criteria-and-working-methods-201902/>
- REF (2020). *Guidance on Revisions to REF2021*. Available at: <https://ref.ac.uk/publications-and-reports/guidance-on-revisions-to-ref-2021/>
- REF (2022a). *Results and Submissions*. Available at: <https://ref.ac.uk/>
- REF (2022b). *Main Panel Overview Reports*. Available at: <https://www.ref.ac.uk/publications-and-reports/main-panel-overview-reports/>
- Sivertsen, G. (2017). 'Unique, but still best practice? The Research Excellence Framework (REF) from an international perspective', *Palgrave Communications*, **3**, art. 17078.
- Technopolis (2015). *REF Accountability Review: Costs, Benefits and Burden*. Report by Technopolis to the four UK higher education funding bodies.
- Thorpe, A., R. Craig, D. Tourish, G. Hadikin and S. Batistic (2018). 'Environment submissions in the UK's Research Excellence Framework (2014)', *British Journal of Management*, **29**, pp. 571–587.
- Walker, J. T., E. Fenton, A. Salter and R. Salandra (2019). 'What influences business academics' use of the Association of Business Schools (ABS) list? Evidence from a survey of UK academics', *British Journal of Management*, **30**, pp. 730–747.
- Wilsdon, J. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. Available at: https://fapesp.br/avaliacao/manuais/2015_metric_tide.pdf
- Zacharewicz, T., B. Lepori, E. Reale and K. Jonkers (2019). 'Performance-based research funding in EU member states—a comparative assessment', *Science and Public Policy*, **46**, pp. 105–15.

Robert Blackburn is Professor of Entrepreneurship at the University of Liverpool Management School and Director of the Brett Centre for Entrepreneurship. He has served on three assessment exercises and was chair of Sub-Panel 17, Business and Management, REF2021. He is a Fellow of the Chartered Association of Business Schools and the Academy of Social Sciences.

Sally Dibb is Professor of Marketing and Society at Coventry University. She is a Fellow of the CIM, recipient of the Academy of Marketing lifetime achievement award and a Fellow of the BAM Peer Review College. Sally served as a sub-panel member on REF2014 and was an international panel member of the 2018 Norwegian national research evaluation. She was co-deputy chair of Sub-Panel 17, Business and Management, REF2021.

Ian Tonks is Professor of Finance at the University of Bristol Business School. He has previously served as a sub-panel member on REF2014; was convenor of the Economics and Business Panels, Hong Kong RAE2020; and was co-deputy chair of Sub-Panel 17, Business and Management, REF2021. He is a Fellow of the Academy of Social Sciences.