

What is a high-quality research environment? Evidence from the UK's research excellence framework

Matthew Inglis ^{1,*}, Elizabeth Gadd² and Elizabeth Stokoe³

¹Centre for Mathematical Cognition, Loughborough University, Loughborough LE11 3TU, United Kingdom

²Research & Innovation Office, Loughborough University, Loughborough LE11 3TU, United Kingdom

³Department of Psychological and Behavioural Science, The London School of Economics and Political Science, London WC2A 2AE, United Kingdom

*Corresponding author. Email: m.j.inglis@lboro.ac.uk

Abstract

As part of the UK university sector's performance-related research funding model, the 'REF' (Research Excellence Framework), each discipline-derived 'Unit of Assessment' must submit a statement to provide information about their environment, culture, and strategy for enabling research and impact. Our aim in this paper is to identify the topics on which these statements focus, and how topic variation predicts funding-relevant research environment quality profiles. Using latent Dirichlet allocation topic modelling, we analysed all 1888 disciplinary 'unit-level' environment statements from REF2021. Our model identified eight topics which collectively predicted a surprisingly large proportion—58.9%—of the variance in units' environment scores, indicating that the way in which statements were written contributed substantially to the perceived quality of a unit's research environment. Assessing research environments will increase in importance in the next REF exercise and the insights found through our analysis may support reflection and discussion about what it means to have a high-quality research environment.

Keywords: research evaluation; research policy; research excellence framework; universities; research culture; research environment.

The research excellence framework

For the past four decades, higher education institutions in the UK have been subject to evaluations of their research by the higher education funding councils. The first evaluation, the 'Research Selectivity Exercise (RSE)' took place in 1986 (for a history, see [Bence and Oppenheim 2005](#)). Over the years, and with six assessments between 1986–08, the RSE evolved into the Research Assessment Exercise (RAE) and then, in 2014, into the 'Research Excellence Framework' (REF). For each evaluation, university disciplines and fields of study were divided into 'Units of Assessment' (UoAs). The most recent assessment took place in 2021, with results published in 2022.

As well as name changes, the requirements for submissions have evolved (see [Marques et al. 2017](#)), from the "quick and dirty" ([Jones and Sizer 1990](#)) approach taken in 1986 through to including/excluding particular categories of staff; changing the minimum/maximum numbers of publications per individual; the introduction of research environment statements (RAE 1996), and the introduction of impact case studies (REF 2014). Two constants about RSE/RAE/REF remain: the original principle of peer assessment, despite the rise of publication metrics in other domains, and the use of the results to distribute government funding.

The RSE represented the creation of "the first and most highly institutionalised research evaluation system worldwide" ([Marques et al., 2017: 822](#)). Since then, the RAE/REF has been widely discussed and used as a model for other countries (e.g. [Geuna and Martin 2003](#)) or resisted and rejected (e.g. [Swedish Government 2016](#)), but rarely adopted wholesale in countries internationally ([French, Massy and Young 2001](#); for overviews,

see [Sivertsen 2017](#); [Thomas et al. 2020](#); [Pinar and Horne 2022](#)). Either way, the discourse of the RAE/REF reaches far beyond the UK.

Analysing the research excellence framework

Unsurprisingly, the RAE/REF has been scrutinized in terms of (i) critiques of the politics and methodologies that underpin the process and, (ii) quantitative and qualitative analyses of submissions, assessment processes, and results themselves. The former comprises a literature too vast to cover substantially here, but includes criticisms of the trend towards a competitive, neoliberal, and commodified higher education system (e.g. [Fairclough, 1995](#); [Brown and Carasso, 2013](#)), of the impact of assessment on individual disciplines and inter-disciplinarity (e.g. [Pardo-Guerra 2022](#)), and of unintended consequences (for overviews, see [Gillies 2008](#); [Brassington 2022](#); [Pinar and Horne 2022](#); [Watermeyer and Derrick 2022](#)). The RAE/REF has driven both policy and debate in UK higher education, with a series of consultations, evaluations, recommendations, and iterated processes ([Manville et al. 2015](#); [Curry, Gadd and Wilsdon 2022](#)).

Another approach to evaluating and critiquing the RAE/REF focuses on the actual content of HEIs' submissions to RAE/REF, using both quantitative and qualitative methods. Several of these studies have used similar text-mining or topic modelling and related methods to those used in this paper. Perhaps because of the increasing significance of research impact over the past decade ([Derrick and Samuel 2016](#); [Kellard and Śliwa 2016](#); [Sutton 2020](#); [Jensen, Wong and Reed 2022](#)), several studies have scrutinized the content and composition of case studies. For instance, a report commissioned by the

UK research funding councils (King's College London and Digital Science 2015) used text-mining and qualitative analysis to provide an initial assessment of all REF2014 impact case studies, making observations about the diverse range of impacts, their underpinning research, and their global reach (see also Terämä et al. 2016). Reichard et al. (2020) conducted two studies using qualitative thematic and quantitative linguistic analysis of REF2014 impact case studies to identify what individual words and phrases were associated with high and low scores. They identified numerous “lexical bundles” associated with lower (e.g. “involved in”, “has been disseminated”, “the event”) and higher (e.g. “the government’s”, “in the UK”) scores, the former associated with describing activities and pathways to impact and the latter evidence of significant and far-reaching impact itself.

Within the wider literature on REF methodology and submissions, the least scrutinized aspect is the environment statement (Thorpe et al. 2018a). We now turn to discuss what we know and do not know about REF environment statements, starting by describing the submission requirements for the 2021 assessment.

The REF environment statement

The environment statement(s) and their submission requirements

As noted above, an environment statement was introduced relatively early into the UK RAE/REF cycle. In 2021, two main changes occurred: the introduction of a pilot “Institutional-level environment statement”, and the removal of an ‘impact template’ from REF2014 and the incorporation of ‘research and impact’ as one environment statement in 2021.

The aim of REF2021 unit-level environment statements was to provide assessable information about each UoA’s “environment for research and enabling impact” (Guidance on Submissions, p. 82) and especially its “vitality” (“the extent to which a unit supports a thriving and inclusive research culture for all staff and research students, that is based on a clearly articulated strategy for research and enabling its impact, is engaged with the national and international research and user communities and is able to attract excellent post-graduate and postdoctoral researchers”) and “sustainability” (“the extent to which the research environment ensures the future health, diversity, wellbeing and wider contribution of the unit and the discipline(s), including investment in people and in infrastructure”) (Panel Criteria and Working Methods, p. 58). For REF 2021, detailed guidance notes and a template were provided to structure the information in four sections: (1) “unit context, research and impact strategy”; (2) “people, including: staffing strategy and staff development, research students, equality and diversity”; (3) “income, infrastructure and facilities”, and (4) “collaboration and contribution to the research base, economy and society.” The permitted length of the statement varied according to the number of staff in a UoA, from 8,000 (for a submission comprising 1–19.99 FTE) to 12,000 words (plus 800 further words per additional 20 FTE). The environment statement was worth 15% of the funding allocation. For Main Panels A, B, and C, each of the four subsections attracted equal weighting; in Main Panel D, sections (1) and (4) attracted 25%, the ‘People’ section 30%, and ‘income, infrastructure and facilities’ 20%.

Analyses of environment statements

To the best of our knowledge, no close analyses of REF2021 environment statements have yet been published, apart from Manville et al.’s (2021) real-time study of REF as it happened, including focus group research on academic and professional service staff experiences of completing them. However, research has been conducted on REF2014 environment statements. For example, Matthews and Kotzee (2022) analysed both REF (2014 submission) and TEF (Teaching Excellence Framework, 2017 submission) documentation with the aim of investigating links between research and teaching. They found that the term “research-led”, analysed in the context of its collocates, was often used in connection with teaching, and argued that “according to what universities themselves write in institutional texts, teaching and research are not always in a mutually beneficial entanglement, but often rather a one-way relationship in which research expertise and institutional prestige are used to bolster claims of teaching excellence” (p. 578).

Mellors-Bourne, Metcalfe and Gill (2017) also used text-mining to assess the level of engagement with equality and diversity in the ‘People’ section of REF2014 environment statements. This included evidence of participation in schemes such as ‘Athena Swan’ and Stonewall’s, and the relative frequencies of words pertaining to the topic: ‘equality’, ‘diversity’, ‘Athena’, ‘gender’, and ‘ethnicity’. Mellors-Bourne et al. found that statements focused predominantly on gender; that “the word ‘equality’ was used on average between once and twice within each environment statement” (p. 2), and that the Main Panels A and B (the STEM disciplines) mentioned ‘Athena’ more than twice as much as Main Panels C and D. The researchers also found “[e]vidence suggesting a positive relationship between REF research environment sub-profiles (scores) and reference to key E&D terms within submissions, overall and at the level of the main panels” (p. 2). In REF2021, the focus on equality, diversity and inclusion in the guidance notes extended well beyond the ‘People’ section, presumably to encourage HEIs to demonstrate how EDI strategies and outcomes were embedded in all areas of research and impact.

Thorpe et al. (2018a, 2018b) used computer-assisted text analysis to scrutinize the language content and ‘tone’ of REF2014 environment statements in business and management schools. They sought to understand whether the way environment statements were written differed between high and low scoring universities. They found that higher-ranked universities used less passive voice, were more coherent, adopted “a ‘finished article’ discourse rather than a ‘we are developing’ discourse”, cited “specifics rather than generalities”, and were more self-referential (Thorpe et al., 2018a, p. 582). In terms of tone, the authors found that, perhaps counterintuitively, higher-scoring statements scored lower in terms of ‘activity’ tone “that evokes a ‘safe’, ‘staid’, ‘orthodox’, ‘conservative’, and ‘settled’ environment that is not disturbed (unduly at least) by reform, disruption, or major staff turnover” (Thorpe et al. 2018b: 60). The authors concluded that “low-ranked universities could have achieved higher scores by reflecting on particular areas of word choice and the potential effects of those choices on assessors” (Thorpe et al. 2018b: 53).

Like Reichard et al.’s (2020) analysis of impact case studies, Thorpe et al. (2018a, 2018b) revealed the importance of language as well as content in the production of environment

statements. They concluded that “the accompanying narrative played an important role in determining REF2014 environment scores” (2018a: 572). Furthermore, in contrast to impact case studies, which included corroborating evidence as part of submission, “little supporting evidence was required in environment submissions”, meaning that “there is potential for writing quality to have an even larger effect in environment submissions, and for HEIs to use language-related techniques to manage their image” (Thorpe et al. 2018a: 574).

In June 2023 the ‘Future Research Assessment Programme’ (FRAP) published their Initial Decisions on REF 2028 (Joint UK HE Funding Bodies 2023), although it has since been announced that the exercise will be delayed until 2029. Despite the Institution-Level Environment Panel Pilot Panel (Research England 2022) recommending the removal of unit-level environment statements and focusing instead on institution-level statements, the Initial Decisions propose the retention and assessment of both. In the 2029 exercise, ‘Environment Statements’ will be broadened to become ‘People, Culture & Environment (PCE) Statements’ with a concomitant increase in weighting from 15% to 25%. The Decisions state that “the collection of evidence for the people, culture and environment element will move towards a more tightly defined, questionnaire-style template that will create greater consistency across submissions and focus on demonstrable outcomes (Joint UK HE Funding Bodies 2023)”. At the time of writing this is yet to be developed and it is unclear what proportion of the submission will take a narrative format and what proportion will comprise data and evidence. However, given the qualitative nature of the dimensions being assessed, it is likely that there will be a significant narrative element. Furthermore, given the increased weighting of PCE, any such element should have an even greater bearing on overall results.

In sum, the focus on environment statements as a subset of research on the UK REF has, to date, been small. It has either been partial (e.g. has analysed just one UoA (Thorpe et al. 2018a, 2018b)) or not focused on the most recent exercise. The aim of this paper is to investigate whether thematic patterns may be identified in the 2021 submissions, whether such patterns may be correlated with scores, and what this might teach us about crafting future environment narratives. We begin by introducing the method we adopted, latent Dirichlet allocation topic modelling.

Topic modelling the research excellence framework

Topic modelling is a computation method that seeks to analyse the content of many texts by identifying a small number of semantically connected themes or *topics* (Blei, Ng and Jordan 2003). The aim is to take a collection of unstructured texts and identify the topics they cover by studying their words. For example, if a document uses the words ‘water’, ‘sand’, and ‘swimming’ with an unusually high frequency, this may constitute evidence that the document is about beaches. One way to understand the topic modelling approach is to think about how to create documents from a pre-defined set of topics, defined to be probability distributions over words. For instance, our beach topic might assign very high probabilities to ‘water’, ‘sand’ and ‘swimming’, medium probabilities to ‘inflatable’, ‘spade’ and ‘picnic’, and low

probabilities to ‘dioxide’, ‘carpentry’ and ‘vener’. Similarly, we might define a topic about Greece (which might perhaps have a high probability associated with the words ‘Greek’, ‘Athens’, ‘souvlaki’ and so on). If we wanted to create a new document about beach holidays in Greece, we might choose 30% of the new documents words to be from the beach topic, 30% from the Greece topic, perhaps 30% from a travel topic, and 10% from other topics. A document about Greek beach holidays can then be created by simply sampling words from each topic using the appropriate probabilities. This method makes two simplifying assumptions. First, the ‘bag of words’ model of text is adopted by ignoring word order; and second, so-called ‘stop words’ (words such as ‘the’, which is topic independent) are ignored.

The topic modelling approach can be thought of as carrying out this document construction process in reverse. We start with a large collection of texts, assume that they were created in this way, and then computationally identify the topics that would have been most likely to lead to these documents using a latent Dirichlet allocation (LDA) algorithm (Blei, Ng and Jordan 2003; Grimmer and Stewart 2013). Topic modelling adopts a grounded theory mentality: the analyst has no preconceived ideas about what topics will be identified, instead topics/themes emerge from the analysis process. Once topics have been identified, the semantic content of each document can be analysed by studying the topic composition of each document. For instance, we may find that Document 1 contains 6% of words from Topic 1, 30% from Topic 2, 0% from Topic 3, and so on.

We downloaded all 1888 unit-level environment statements (a total of 18.0 m words) from the REF website. These were converted from pdf into plain text using the UNIX `pdftotext` command (Poppler 2022). We used MALLET (version 2.0.8RC2), a UNIX topic modelling tool (McCallum 2002), to calculate possible models, using MALLET’s default list of stop words.

Topic modelling requires that one specifies how many topics the LDA algorithm should identify. By making different choices researchers can specify the granularity of their analysis. We adopted the *perplexity* approach to decide on the number of topics. Each model can be assigned a perplexity, which is analogous to a model fit (Blei, Ng and Jordan 2003). Perplexity can be calculated by fitting a model with a specified number of topics to a subset of the documents, and then assessing its fit to the remaining documents. One can always reduce the perplexity (or increase the fit) by fitting a model with a larger number of topics, although at some point the benefit of doing so will be offset by the increased difficulty of interpretation. Jacobi, van Atteveldt and Welbers (2016) suggested using a method similar to the scree test often used in factor analyses: by calculating the perplexity of models with a range of different topic numbers, it is possible to determine if there is a point at which the benefit, in terms of reduced perplexity, of increasing the number of topics appears to level off.

We split the environment statements into a training corpus (80% of statements) and a testing corpus (20% of statements), fitted models with 10, 20, 30, ..., 100 topics to the training corpus, and calculated the associated perplexities using the testing corpus. These perplexity figures are shown in Figure 1. We then fitted a piecewise linear regression to these points, which suggested that the ‘elbow’ of the graph

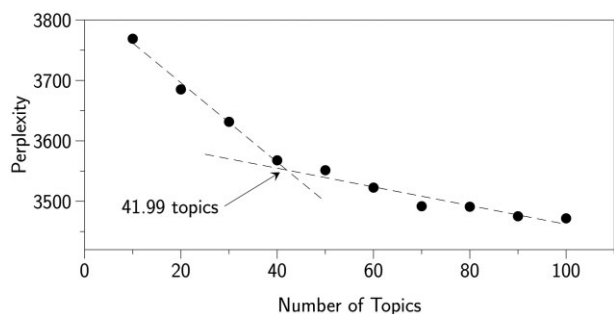


Figure 1. Perplexities associated with models with 10, 20, 30, ..., 100 topics. The dotted lines show a piecewise linear regression line of best fit.

appeared at 41.99 topics. We therefore selected 42 topics for our main analysis.

Topic modelling has an important advantage over more traditional qualitative analytical techniques in that it is extremely inclusive. Given that 1888 unit-level environment statements were returned to REF2021, containing ~ 18 m words, it would have been impractical for a human analyst to read and analyse each statement. However, topic modelling is not purely quantitative: the LDA algorithm identifies topics which must then be interpreted. One common approach to this task is to conduct careful qualitative analyses of documents that contain a high proportion of words from each topic. We return to this issue later in the paper.

The topics

The 42 topics identified are shown in Table 1. The table shows the characteristic words associated with each topic, the statement with the highest proportion of words from that topic, and the label we gave the topic. These labels were based on our interpretations of studying the characteristic words, the statements with particularly high proportions of words from the topic, and the statements with particularly low proportions of words from the topic.

Of the 42 topics, 28 were disciplinary specific. For example, Topic 42 was characterised by words including ‘clinical’, ‘cancer’, ‘medicine’, ‘disease’, ‘MRC’ and ‘NHS’, and the top 10 environment statements in terms of the proportion of words with this topic were all returned to the Clinical Medicine panel. We therefore labelled this topic ‘Clinical Medicine’. In some cases our model combined two or more disciplines. For instance, Topic 2 was characterised by the words ‘philosophy’, ‘religion’, ‘theology’, ‘religious’ and ‘ethics’. Of the top 20 statements with high proportions of words from this topic, 9 were returned to the Philosophy panel and 11 to the Theology and Religious Studies panel. We used the label ‘Philosophy and Religion’.

There were five geographical topics. For instance, Topic 8 was characterised by the words ‘Liverpool’, ‘Leeds’, ‘Manchester’, ‘York’, ‘Sheffield’ and ‘Yorkshire’, all cities/regions in the north of England, and the statements with the highest proportion of words from this topic were from northern universities. There were geographical topics associated with the North, the West Country, Scotland, London and Wales.

We also found a topic, Topic 25, that was used by institutions that organise their academic work through constituent colleges. The topic was characterised by words including

‘university’, ‘faculty’, ‘centre’, ‘college’ and ‘institute’, as well as geographical terms that referenced multi-college universities (‘London’, ‘Oxford’, and ‘Cambridge’). Of the 25 statements with the highest proportions from this topic, 23 were returned by constituent faculties or colleges of the University of Oxford, the University of Cambridge or the University of London. The exceptions were statements from Kingston University (an institution based in London) and Oxford Brookes University (an institution based in Oxford).

Of most interest for our purposes are the remaining eight topics. To identify appropriate labels we followed a similar process, separately for each topic. First, we studied the topic’s characterising words. Second, we read the five environment statements with the highest proportion of words from the topic, and the five environment statements with the lowest proportion of words from the topic. Finally, we conducted concordance analyses to identify how characterising words were used in statements with high and low proportions of words from the topic. This involved using a keyword in context (KWIC) tool from a traditional corpus linguistics package (Anthony 2022). For instance, if ‘faculty’ was a characteristic word for a topic, we would find every occurrence of ‘faculty’ in the five statements with the highest proportion of words from the topic and read the surrounding context to identify how the word was typically being used. We would then do the same for the five statements with the lowest proportion of words from the topic. To illustrate our approach, we first discuss the reasoning behind our naming of Topic 16 in some detail, and then discuss each of the remaining seven topics in turn. Note that the online data associated with this article includes the topic weightings derived from our model for each topic and each environment statement submitted to REF2021, so interested readers can independently verify our analyses and assess for themselves the topic names’ appropriateness.

Although all REF environment statements are publicly available from the Research England website, we opted to redact individuals’ names in the quotes reported below as they are not relevant to the research questions we asked. We have, however, not anonymised at the institution or department level, as these may assist readers interpret the topics.

Topic 16—Immature Research Environment

The proportion of words from Topic 16 used by statements varied from 0.0001 (Imperial College’s Mathematical Sciences statement) to 0.362 (Bedfordshire’s Business and Management Studies statement). The topic was characterised by words such as ‘research’, ‘university’, ‘staff’, ‘international’, ‘REF’, ‘member’, and ‘members’ (see Table 1). Unlike with some of the other topics discussed below, we did not find these words very insightful for determining the semantic content of the topic.

Our next step was to carefully read the five statements with the highest proportion of words from the topic and the five statements with the lowest proportion of words from the topic. This revealed that the topic-defining words seemed to be being used in characteristic ways by those statements with a high proportion of words from the topic. Specifically, Topic 16 was characterised by descriptions of how the units were trying to encourage staff to engage in research. For example, the Wrexham Glyndŵr University Computer Science and Informatics statement (Topic 16 proportion 0.327) noted that “Data from October 2020 indicates that 38% of the 13 members of academic staff associated with UoA11 [the

Table 1. Descriptive statistics, averaged across each of the four quarters of the experiment, for each of the four indices under consideration

Topic name	Topic name	Characteristic words (top 20)	Statement with the highest proportion of words from the topic (University—Unit)
1	Education	education educational learning practice teaching schools professional research teacher teachers higher doctoral social children colleagues policy development language e.g. international	Manchester Metropolitan University—Education
2	Philosophy and Religion	philosophy religion theology religious ethics philosophical studies keele humanities project society ahrc department mind interdisciplinary church arts public science conference	Leeds Trinity University—Theology and Religious Studies
3	Chemistry	chemistry materials chemical epsrc facilities rsc molecular energy industrial equipment industry group synthesis catalysis analytical facility phd nmr chem spectroscopy	University of Edinburgh—Chemistry
4	Internal Structure of Research Units	unit unit's gbp ref faculty pgrs smith supported ics section fte sussex submitted institutional open themes theme northumbria aru uoc	University of Cumbria—Business and Management Studies
5	Sociology, Communication and Culture	social research international work studies policy global media esrc impact public colleagues centre digital political project gender network development culture	Brunel University London—Sociology
6	Law	law legal rights justice colleagues international human school criminal european court policy review committee school's scholars society academics centre criminology	University of Southampton—Law
7	Career Development and EDI	research staff support impact students funding training including university applications access ref open phd external annual academic career members diversity	Leeds Arts University—Music, Drama, Dance, Performing Arts, Film and Screen Studies
8	The North	liverpool leeds manchester york sheffield yorkshire lincoln pgr hull city university external chester uol postgraduate salford north public ljmu local	York St John University—History
9	Physics	physics quantum stfc group materials science facilities space astronomy epsrc matter particle astrophysics technology facility imaging nuclear international laboratory computing	University of Leicester—Physics
10	Health	health nihr clinical uoa trials data public global care primary disease dental covid trial epidemiology mrc unit diseases group oral	London School of Tropical Medicine—Clinical Medicine
11	Politics	politics political international studies policy security european colleagues relations conflict university public esrc government british polis theory peace global foreign	University of Hull—Politics and International Studies
12	Engineering	engineering energy materials systems technology manufacturing epsrc industry industrial technologies facilities advanced design innovation group laboratory power modelling sustainable control	Cranfield University—Engineering
13	Psychology	psychology neuroscience e.g health mental science cognitive psychological social brain group cognition behaviour clinical esrc human behavioural facilities lab experimental	University of Glasgow—Psychology, Psychiatry and Neuroscience
14	Classics and Languages	ireland classics irish northern qub classical ancient ulster phd belfast greek queen's roman cycle unit project reception postgraduate conflict e.g	University of Ulster—Modern Languages and Linguistics
15	West Country	bristol faculty exeter iles bath statement plymouth west university uwe uob south southampton dmu aston college group pgr associate cornwall	University of Bristol—Management and Business Studies
16	Immature Research Environment	research university staff international ref students conference development journal pgr phd member funding project support external unit environment members income	University of Bedfordshire—Management and Business Studies
17	Scotland	scottish scotland edinburgh glasgow aberdeen government dundee andrews phd society studies graduate exchange centre stirling knowledge uws review uofg strathclyde	University of the Highlands and Islands—Modern Languages and Linguistics
18	Staff Ways of Working	school work ref colleagues period students research university group environment teaching areas current members grant part number range strategy years	University of East Anglia—Law

(continued)

Table 1. (continued)

Topic name	Topic name	Characteristic words (top 20)	Statement with the highest proportion of words from the topic (University—Unit)
19	Economics	department departmental economics department's research economic phd policy departments faculty students lse journal members impact international durham bank review financial	University College London—Economics and Econometrics
20	Business and Management	business management school innovation finance journal economics marketing financial accounting economic social entrepreneurship policy international faculty leadership sustainable centres government	University of Derby—Business and Management Studies
21	Geography and Environment	environmental climate nerc marine change science geography earth water environment natural management group global energy e.g society staff conservation carbon	Heriot Watt University—Earth Systems and Environmental Sciences
22	Computer Science and Informatics	systems data computer computing science security digital ieee technology software industry epsrc learning engineering phd technologies group intelligence robotics project	University of Salford—Computer Science and Informatics
23	London	ucl studies anthropology soas ucl's east department asia global staff birkbeck africa china south african departmental students london grant middle	School of Oriental and African Studies—Communication, Culture and Media Studies
24	Biological Sciences	biology cell bbsrc uea biological molecular wellcome nature uoa sciences phd evolution disease ucl plant facility biosciences e.g researchers facilities	University College London—Biological Sciences
25	Collegiate Structures	research university faculty students centre public college environment institute oxford unit-level london researchers page ref studies template cambridge collaboration funding	University of Cambridge—Modern Languages and Linguistics
26	English	english literature language studies writing creative literary humanities linguistics languages poetry culture cultural translation work ahrc project modern arts colleagues	University of Essex—English Language and Literature
27	Social Work and Social Policy	social work policy justice crime sociology police criminology health practice violence care policing people criminal children child young abuse centre	University of East Anglia—Social Work and Social Policy
28	REF-Focused Research Strategy	uoa ref uoa's section members programme college uclan open uoas interdisciplinary rke level university's role cycle external birmingham support submitted	University of Winchester—History
29	Sport and Exercise	sport exercise sports physical health performance activity e.g science football sciences physiology psychology group tourism nutrition staff equipment ref laboratory	Hartpury University—Sport and Exercise Sciences, Leisure and Tourism
30	Exemplification of Strategy and Processes	research pgrs impact ref pgr e.g international including support funding staff training grant edi interdisciplinary awards ecrs national supported school	University of Nottingham—Geography and Environmental Studies
31	Archaeology	archaeology heritage archaeological project e.g museum projects human conservation ahrc landscape facilities cultural staff analysis science national digital historic department	University of Aberdeen—Archaeology
32	Public Health	health care nhs research nihr public clinical mental group practice social policy healthcare national people nursing service dementia services medical	The Metanoia Institute—Psychology, Psychiatry and Neuroscience
33	History	history historical heritage colleagues project museum modern humanities british ahrc war studies public national medieval library historians culture society archives	University of Central Lancaster—History
34	Industry Partners and Funding	ref staff science centre including period international award environment industry society data e.g institute unit-level academic training page awards include	University of Bristol—Chemistry
35	Mathematics	mathematics mathematical theory statistics group epsrc analysis sciences phd physics applied data statistical modelling science applications groups geometry institute lms	University of Chester—Mathematical Sciences

(continued)

Table 1. (continued)

Topic name	Topic name	Characteristic words (top 20)	Statement with the highest proportion of words from the topic (University—Unit)
36	Music, Drama, Dance, Performing Arts, Film and Screen Studies	music film arts performance theatre media creative dance cultural practice festival work digital projects ahrc project studies sound drama screen	Canterbury Christ Church University—Music, Drama, Dance, Performing Arts, Film and Screen Studies
37	Architecture, Built Environment and Planning	urban architecture environment energy built design construction planning housing projects sustainable building cities project management transport industry ntu architectural policy	Edinburgh Napier University—Architecture, Built Environment and Planning
38	Wales	wales welsh cardiff swansea government bangor university national cymru researchers centre aberystwyth usw williams period european project society ref coe	University of Wales Trinity St David/ Prifysgol Cymru Y Drindod Dewi Sant—Modern Languages and Linguistics
39	Agriculture, Food and Veterinary Sciences	food animal health agriculture bbsrc systems agricultural plant veterinary production industry development sustainable policy nutrition welfare global science environmental facilities	University of Edinburgh—Agriculture, Food and Veterinary Sciences
40	Early Career Researcher (ECR) Development	research development support researchers impact environment including strategy work template unit-level engagement page strategic develop funding collaboration programme working key	Canterbury Christ Church University—Sport and Exercise Sciences, Leisure and Tourism
41	Art and Design	art design arts creative practice cultural museum gallery digital exhibition exhibitions artists visual culture contemporary film ahrc projects heritage staff	Art and Design: History, Practice and Theory
42	Clinical Medicine	clinical cancer research medicine disease uoa mrc imaging centre translational nhs trials cell medical wellcome patients nihr biomedical university drug	University of Newcastle upon Tyne—Clinical Medicine

Computer Science and Informatics unit] have a doctoral qualification” and that “An encouraging sign is that 38% of UoA11 staff are studying towards a doctorate.” Similarly, the Liverpool John Moores University Business and Management Studies statement (T16 proportion 0.339) noted that 31 members of staff “are being supported in their research-related activities, with a view to them being research active in the next assessment period” and that staff were supported by holding seminars, where the invited external speakers were “editors of peer-reviewed journals with high impact factors” in order to assist staff “target publications in highly respected journals”. The Bedfordshire Business and Management Studies statement (T16 proportion 0.362) emphasised that “Staff members are strongly encouraged to attend international conferences and present their research results” and that their staff “are allocated dedicated research time as part of their workload”. In contrast, the statements with low frequencies of words from Topic 16 seemed to take for granted that their academic staff had doctorates and routinely conducted research.

Next, we conducted concordance analyses comparing how the topic’s defining words were used in the five statements with a high proportion of words from Topic 16 with how they were used in the five statements with a low proportion of words from the topic. For instance, we compared how these statements used the word “research”. In the five high statements there were 19 uses of “research active” (e.g. “continued to be research active”, “support to be research active”, “increased the number of research active staff”, “sought to retain research active staff”, “staff on the cusp of being research-active”) compared to just one in the five low statements, which appeared in a subheading in the University

College London’s Law statement (“2.1 Research-active staff and output selection profile”).

As another example, across the five high Topic 16 statements there were 82 instances of “conference”, including numerous examples of conferences that had been attended by staff from these units. In contrast, this word appeared only 29 times in the five low Topic 16 statements and tended to be used as an illustration of a wider point. For example, in the Cambridge philosophy statement (T16 topic proportion 0.00003), the organisation of the Cambridge Platonism conference was given as an example of the unit’s interdisciplinary research (the conference was jointly organised with the Cambridge Faculty of Divinity).

To give one final example, there was also a difference in the way the high- and low-T16-proportion statements used the word “journal”. The five high-T16-proportion statements contained 37 instances of this word. In some cases, these were examples of how members of the unit had written research articles, e.g. Newman University’s Sport and Exercise Sciences, Leisure and Tourism statement (T16 proportion 0.360) noted that “Visiting Professor [anonymised] has produced a manuscript currently in review in the European Respiratory Journal Open”. In other cases there were lists of interactions with journals: Wrexham Glyndŵr University’s Computer Science and Informatics statement explained how one colleague was “on the review panel for a further 6 journals”. In contrast, the five low-T16-proportion statements had only 13 instances of the word ‘journal’. These tended to be examples of how the unit was contributing to the wider academic community. For instance, the University College London law environment statement (T16 proportion 0.0003) discussed how they were progressing towards an open research environment and exemplified this by noting how one

member of the unit had “founded *Europe and the World: A Law Review* as a fully peer-reviewed OA [open access] journal”.

In sum, we concluded that those statements which had a high proportion of words from Topic 16 tended to spend a large proportion of their statement discussing how they were attempting to encourage or support routine research activities. These kinds of discussions were absent from those statements with a low proportion of words from this topic. We therefore named this topic “Immature Research Environment”.

Topic 4—Internal Structure of Research Units

Topic 4 was characterised by the high use of words such as ‘unit’, ‘unit’s’, ‘faculty’, ‘section’, ‘themes’, ‘theme’ and ‘institutional’. The proportion of words from this topic ranged from 0.000 (the University of Edinburgh’s Clinical Medicine statement) to 0.150 (the University of Cumbria’s Business and Management Studies statement). The topic tended to be characterised by detailed descriptions of the internal structure of the units. For instance, the University of Cumbria’s Business and Management Studies statement (T4 proportion 0.150) devoted 1.5 pages of their statement to the “unit context and structure” which noted how, during the assessment period, they had created a new institute and developed three new research themes. Similarly, the University of Winchester’s English Language and Literature statement (T4 proportion 0.137) spent just over a page discussing their unit context and structure, noting how the unit was situated within the University’s department and faculty structure, how it contained a research centre, and how the Centre interacted with other centres across the University. We named this topic “Internal Structure of Research Units”.

Topic 7—Career Development and EDI

Topic 7 was characterised by words such as ‘staff’, ‘support’, ‘training’, ‘including’, ‘access’, ‘career’ and ‘diversity’. The proportion of words from this topic ranged from 0.004 (The Royal Agricultural University’s Agriculture, Food and Veterinary Sciences statement) to 0.399 (Leeds Arts University’s Music, Drama, Dance, Performing Arts, Film and Screen Studies statement). Statements with a high proportion of words from Topic 7 tended to have long sections that discussed how the unit supported staff and student development, and about their equality and diversity processes. For instance, the University of Nottingham’s Politics and International Studies statement (T7 proportion 0.290) included careful statistical analyses of their gender balance at different career stages, as well as analyses of their staff by ethnicity, disability and age profiles. The statement then went on to discuss how these analyses informed “EDI-focused improvements”. For instance, in response to “too few members from underrepresented groups in leadership roles” the statement noted how the unit had reconstituted its EDI committee and “increased leadership by women in major committees”. In contrast, the Royal Agricultural College’s Agriculture, Food and Veterinary Sciences statement (T7 proportion 0.004) devoted just 50 words to EDI issues, and only used the word ‘diversity’ in the context of their research on “the global distribution of earthworm diversity”. We named this topic “Career Development and EDI”.

Topic 18—Staff Ways of Working

Topic 18 was characterised by words such as ‘work’, ‘school’, ‘colleagues’, ‘teaching’, ‘group’, ‘members’, ‘part’ and ‘years’.

Some statements contained a very low proportion of words from this topic, e.g. Heriot-Watt University’s Architecture, Built Environment and Planning statement (T18 proportion 0.000), whereas others contained a substantial proportion from it, e.g. the University of East Anglia’s Law statement (T18 proportion 0.262). The statements with a high proportion of words from the topic were characterised by many concrete descriptions of staff working practices. For example, the University of Newcastle upon Tyne’s Classics statement (T18 proportion 0.244) described how “Members who have held a substantial administrative role are entitled to an extra semester of research leave”. Similarly, the University of St Andrews’s Economics and Econometrics statement (T18 proportion 0.255) discussed the process by which academic staff can apply for sabbatical leave: “The HoS considers applications in relation to the general workload allocation process and, if there are doubts about the feasibility of accommodating all applications, the HoS consults a panel of senior colleagues.”

In our concordance analysis we noted that ‘work’ was commonly used as a verb in high-T18-proportion statements to describe concrete examples of how the unit operated (“the University granted [anonymised] two years of unpaid leave to work and develop Impact at the European Central Bank”), whereas in low-T18-proportion statements it was often used as a noun (“our work on public health engineering”). We also found that ‘members’ was more often used in high-T18-proportion statements to describe concrete internal activities (“Individual staff members can request travel funding and leave to attend masterclasses and short courses”), whereas in low-T18-proportion statements it was typically used to describe esteem activities (“Our researchers also chaired or have served as members of important grant panels”). We named this topic “Staff Ways of Working”.

Topic 28—REF-Focused Research Strategy

Topic 28 was characterised by words such as ‘UoA’, ‘REF’, ‘UoA’s’, ‘UoAs’, ‘section’, ‘cycle’ and ‘submitted’. The proportion of words from this topic varied from 0.000 (the University of Edinburgh’s Clinical Medicine statement) to 0.126 (the University of Winchester’s History statement). Statements with a high proportion of words from Topic 28 tended to use REF terminology to describe their research environment. For example, they might characterise their internal structure in terms of UoAs or ‘units’ rather than departments, centres or institutes. For instance, the University of Worcester’s English Language and Literature statement (T28 proportion 0.126) described “The Unit’s strategic research objectives”, “the unit’s impact strategy” and “the unit team”; and the University of Winchester’s History statement (T28 proportion 0.126) described how “the UoA had a devolved budget”, the existence of a “UoA working group” and “the strategic aims of the UoA over the cycle”. In contrast, the joint engineering statement from the University of Edinburgh and Heriot-Watt University (T28 proportion 0.000) contained no instances of ‘UoA’, and only used “unit” in the generic text used in the page header (“unit-level environment template (REF5b)”). Instead, they described how their research was organised into “cross-cutting organisational themes” and “interdisciplinary global research challenge areas”. Similarly, the University College London education statement (T28 proportion 0.000) discussed how their research was organised into departments and research centres, and only used the word ‘UoA’ in a table reporting

the submission's demographic data. We named this topic "REF-Focused Research Strategy".

Topic 30—Exemplification of Strategy and Processes

Words that characterised Topic 30 included "e.g.", "including", "funding", "supported", "grant" "PGRs", "impact" and "awards". Like Topic 16, it was not immediately obvious to us from studying these words what the topic referred to. However, when we compared the high-T30-proportion statements and the low-T30-proportion statements, we concluded that the topic was capturing an increased use of concrete examples to illustrate strategies and processes. To illustrate, the five statements with the highest proportion of words from Topic 30 made liberal use of "e.g." to give explicit examples of the research strategies being described. For example, the University of Nottingham's Geography and Environmental Studies statement (T30 proportion 0.278) described how they enable and facilitate impact: "the new Institutional Institute of Policy and Engagement has helped fund pump-priming engagement work (e.g. [anonymised]); fund high-level policy relevant talks (e.g. [anonymised] at Asia House and Chatham House) and aid development of policy briefs (e.g. [anonymised] on water management in the Red River, French on indebtedness and financial exclusion)." In their section on open research, they wrote that "The School developed and hosts online, openly accessible maps, including the Blue-Green Cities multiple benefits toolbox ([anonymised]) and the 'black presences and the legacies of slavery and colonialism' online map ([anonymised])". Similarly, the University of Leicester's Communication, Cultural and Media Studies, Library and Information Management statement (T30 proportion 0.266) noted that their "strategy of enabling researcher development in Media focuses on supporting ECRs and mid-career academics, to achieve external funding success. For example, 23 of the 40 awards secured in the REF period were to Assistant Professors."

While low-T30-proportion statements also used exemplification, these tended to be less related to the research strategies and processes described in the statements. For instance, there were 22 instances of "e.g." in Imperial College London's Clinical Medicine statement (T30 proportion 0.000), of which 12 were used in front of lists of journals ("Our reach is demonstrated by publishing in specialist journals (e.g. *Lancet Infect Dis* [11], *Nature Immunology* [7])") and a further 4 were used in front of scientific concepts ("... to reveal mechanisms of cardiovascular disease. e.g. identification of titin variants in health and disease").

We named Topic 30 "Exemplification of Strategy and Processes".

Topic 34—Industry Partners and Funding

Topic 34 was characterised by words such as 'award', 'awards', 'industry', 'society', 'data', and 'international'. The proportion of words from this topic in statements varied from 0.000 (the University of East Anglia's Area Studies statement) to 0.259 (the University of Bristol's chemistry statement). Those statements which had a high proportion of words from Topic 34 devoted considerable space to discussing their industrial partnerships and research funding. For example, the Imperial College London Chemistry statement (T34 proportion 0.252) noted that "Collaborations with industry include GSK and Pfizer", and that "members are involved in industry collaborations e.g. a £3.2M EPSRC BP Prosperity Partnership". The

University of Surrey Physics statement (T34 proportion 0.249) argued that their "world-class research is evidenced by grant awards over the REF period that total more than £19.6 million" and noted that they "work closely with industrial partners". Given this focus, it was unsurprising that the correlation between the proportion of an environment statement made up of words from Topic 34 was strongly correlated with a submission's research funding per FTE, $r = 0.642$, $P < 0.001$. Notably, however, this correlation was much reduced if research income was standardised within each UoA (to $r = 0.275$, $P < 0.001$). In other words, Topic 34 related to overall unstandardised research funding, meaning that statements with a particularly high proportion of Topic 34 words tended to come from highly funded scientific disciplines. Indeed, the mean proportion of words from Topic 34 for statements to Main Panels A (medicine, health and life sciences), B (physical sciences, engineering and mathematics), C (social sciences) and D (arts and humanities) were 0.109, 0.139, 0.044 and 0.020 respectively. In other words, statements from scientific disciplines tended to use more words from Topic 34 than statements from non-scientific disciplines, an observation consistent with our conclusion that the topic concerned industrial partnerships and funding. We named this topic "Industry Partners and Funding".

Topic 40—Early Career Researcher (ECR) Development

The last of our eight general topics was Topic 40. This topic was characterised by words such as 'development', 'develop', 'support', 'research', 'researchers', 'strategy', 'strategic', 'work', 'working' and 'funding'. Of the eight general topics, Topic 40 was the most common: on average environment statements devoted 18.6% of their content to it, although the proportions of words from Topic 40 ranged from 0.038 (Kingston University's Philosophy statement) to 0.403 (Canterbury Christ Church University's Sport and Exercise Sciences, Leisure and Tourism statement). Those statements with a high proportion of words from Topic 40 spoke at length about researcher development, with a particular focus on early career researchers. For instance, the Queen Margaret University Edinburgh Sociology statement (T40 proportion 0.360) discussed how they "support researchers in exploring and preparing for a diversity of careers, for example, through the use of mentors and careers professionals, training, and secondment" and the Solent University Southampton Sport and Exercise Sciences, Leisure and Tourism statement mentioned that a "research mentoring programme organised through [the School Advisory Group for Research] has been implemented to support researchers". The five statements with the highest proportion of words from Topic 40 made 23 references to the Concordat to Support the Career Development of Researchers compared to 4 references in the five statements with the lowest proportion of words from this topic. We named the topic "Early Career Research (ECR) Development".

Predicting environment scores

For our main analysis, we asked whether the eight general topics that environment statements focused on were related to the quality profiles they received. Recall that panels assessed each submission's environment using a five-point scale from 'unclassified' ("an environment that is not conducive to producing research of nationally recognised quality or enabling impact

of reach and significance”) through to ‘4*’ (“an environment that is conducive to producing research of world-leading quality and enabling outstanding impact, in terms of its vitality and sustainability”). Each submission was awarded a ‘quality profile’ based on its environment statement and associated data (discussed below). For instance, the Open University’s submission to the Classics UoA was rated as having an environment where 25% of activity was 4* (world-leading), 50% was 3* (‘internationally excellent’), 25% was 2* (‘recognised internationally’) and 0% was 1* (‘recognised nationally’) or unclassified. For each submission we calculated a grade point average (GPA), which was a simple linear combination of the percentage of each quality level. So the Open University’s Classics submission obtained an environment GPA of 3.0 ($0.25 \times 4 + 0.5 \times 3 + 0.25 \times 2 + 0 \times 1 + 0 \times 0$).

Alongside environment statements, the assessment panels were also provided with additional metrics associated with each submission. These included the full-time equivalent number of staff (FTE) being returned in the submission, the grant income that the unit had received during the assessment period (which could be broken down by source and date), and the number of doctoral degrees that the unit had awarded during the assessment period.

We ran a hierarchical regression predicting each unit’s environment GPA. In the first block we entered each unit’s FTE, their research income per FTE, and the number of doctoral degrees awarded per FTE. Each of these metrics was standardised (using *z* scores) within each UoA to take account of disciplinary norms (for instance, the mean grant income per FTE in the Clinical Medicine unit was £3.7 m compared to £74k in the English Language and Literature unit). In the second block we entered the proportion of each environment statement from the eight general topics discussed above.

The results of this regression are shown in Table 2. Together the environment metrics could explain 47.3% of the variance in environment GPAs. When the topic weightings were added, an additional 21.9% of the variance could be explained, bringing the overall R^2 to 69.1%. Thus the weightings of these eight topics explained significant extra variance in environment GPAs, $F(8, 1870) = 166, P < 0.001$. When the eight topic weightings were used as predictors in the first block (ie before the metrics were entered) they explained 58.9% of the variance in environment GPAs, $F(8, 1873) = 336, P < 0.001$. In sum, the weightings associated with the eight general topics in our topic model predicted a surprisingly large proportion of the variance in submissions’ environment GPAs, indicating that the topics that environment statements focused upon made a substantial contribution to the perceived quality of each submission’s research environment.

As shown in Table 2, of the eight topics, four were significant negative predictors of environment GPA, two were significant positive predictors and two were not significant predictors. Statements that had higher weightings from the Immature Research Environment, Staff Ways of Working, REF-Focused Research Strategy, and ECR Development topics were associated with lower environment GPAs. Statements that had higher weightings from the Exemplification of Strategy and Processes, and Industry Partners and Funding topics were associated with higher environment GPAs.

Because this regression analysis only assessed whether topic weightings were linearly associated with environment

Table 2. A hierarchical regression analysis predicting environment GPA with various metrics (entered in Block 1) and topic weightings from the eight general topics (entered in Block 2)

Predictor	Beta	R^2	ΔR^2
Block 1			
Doctoral degrees per FTE (standardised)	0.212***		
Grant income per FTE (standardised)	0.318***		
FTE (standardised)	0.394***	0.473***	0.473***
Block 2			
Doctoral Degrees per FTE (standardised)	0.094***		
Grant Income per FTE (standardised)	0.214***		
FTE (standardised)	0.201***		
Topic 4—Internal Structure of Research Units	−0.006		
Topic 7—Career Development and EDI	−0.023		
Topic 16—Immature Research Environment	−0.438***		
Topic 18—Staff Ways of Working	−0.057***		
Topic 28—REF-Focused Research Strategy	−0.054***		
Topic 30—Exemplification of Strategy and Processes	0.117***		
Topic 34—Industry Partners and Funding	0.068***		
Topic 40—Early Career Researcher (ECR) Development	−0.112***		
		0.691***	0.219***

* $P < 0.05$.

** $P < 0.01$.

*** $P < 0.001$.

GPAs, we also investigated whether there were nonlinear relationships by inspecting scatterplots of topic weightings against environment GPAs separately for each topic. These are shown in Figure 2, together with cubics of best fit. There appeared to be a clearly nonlinear relationship between topic weighting and environment GPA for Topic 7 Career Development & EDI. Placing little emphasis on this topic was associated with receiving a low environment GPA, but so was placing too much emphasis on it. The cubic of best fit obtained its maximum when 13.4% of the statement was made up of words from Topic 7 (recall that this figure is a percentage of all words in the statement, after stop words have been removed). Environment statements varied substantially in the extent that they discussed Career Development and EDI—the statement with the lowest emphasis on this issue had just 0.4% of its words from the topic, the statement with the highest had 39.9%. But the highest environment GPAs, on average, were obtained by statements where 13–14% of the statement focused on Career Development and EDI.

Topic 40 ECR Development also showed a possibly nonlinear relationship between topic weighting and GPA, although this was less clearly the case than for Topic 7. For statements where between 0% and 20% of their words came from Topic 40 there was a reasonably flat relationship with GPA. But those statements with higher proportions from this topic showed a negative relationship between topic weighting and GPA.

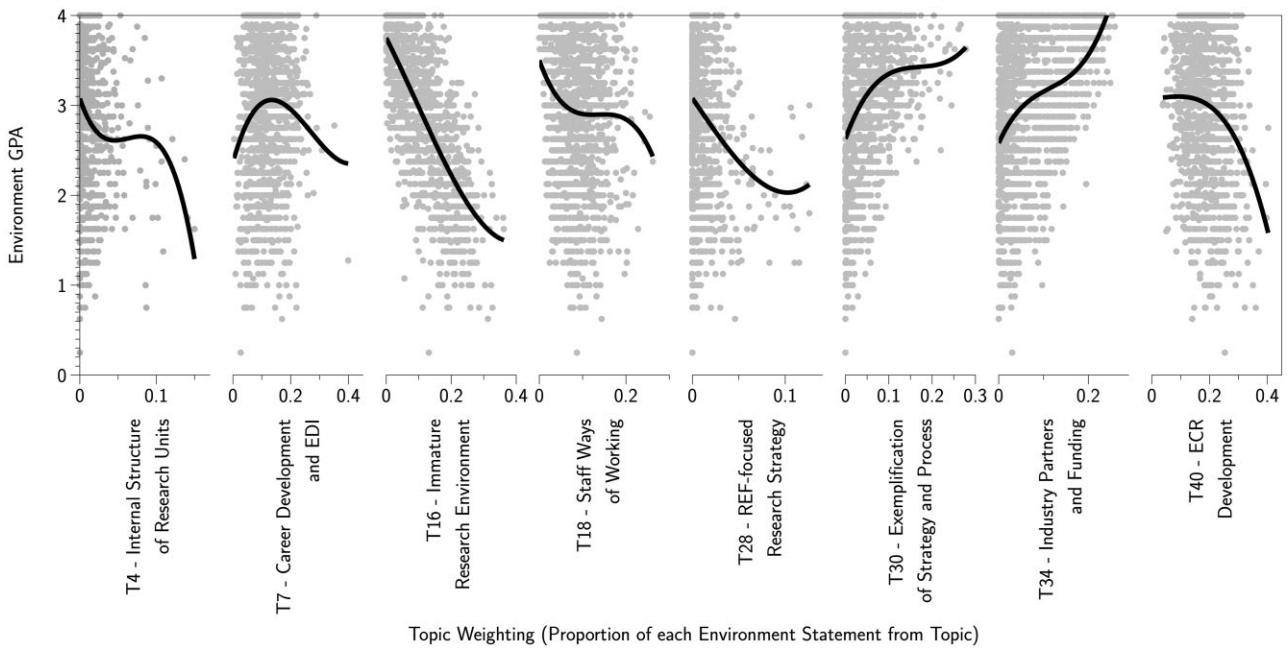


Figure 2. Scatterplots showing topic weightings (proportion of each statement made up of words from the given topic) against environment GPA, separately for the eight general topics. Bold lines are cubics of best fit.

Next, we explored whether environment statements that discussed more discipline-specific issues scored more highly than those which did not. In other words, we asked whether environment statements returned to, say, the Clinical Medicine UoA scored more highly when if they used more words from the Clinical Medicine topic. To investigate this we calculated the correlation between environment GPA and topic weighting for the disciplinary topic, separately for each UoA. These results are shown in Table 3. While a large majority of these correlations were positive, indicating that environment statements that contained more discipline-specific language tended to score higher, there were systematic differences between broad subject areas. The mean correlations between the percentage of discipline-specific language and GPA for Main Panels A (medicine, health and life sciences), B (physical sciences, engineering and mathematics), C (social sciences) and D (arts and humanities) respectively were 0.472, 0.217, 0.172 and 0.093 respectively. For all main panels other than D, these means were significantly greater than zero.

To compare the strength of the association between the extent to which submissions discussed disciplinary issues and their GPAs with the strength of the associations between the eight general topics discussed above and GPAs, we ran a regression analysis on submissions to the Business and Management Studies panel. We chose Business and Management as it was the panel which received the largest number of submissions (108), and so offered the greatest statistical power for an analysis of this kind. In this regression we used the eight general topics, plus the Business and Management topic (Topic 20) to predict environment GPAs. This model is shown in Table 4. Crucially, the Business and Management topic weighting variable was a significant predictor in this model and had a standardised regression coefficient of $\beta = 0.120$, larger than that associated with Exemplification of Strategy and Process (Topic 30, $\beta = 0.102$), and roughly a third the size of Industry Partners and

Table 3. Correlations between environment GPA and disciplinary topic weightings, per UoA (e.g. within the Clinical Medicine UoA, the correlation between environment GPA and weightings on Topic 42 was $r = 0.560$)

Unit of assessment	Disciplinary topic	Correlation
1. Clinical Medicine	42—Clinical Medicine	0.560**
2. Public Health, Health Services and Primary Care	10—Health	0.527**
3. Allied Health Professions, Dentistry, Nursing and Pharmacy	42—Clinical Medicine	0.493***
4. Psychology, Psychiatry and Neuroscience	13—Psychology	0.308**
5. Biological Sciences	24—Biological Sciences	0.571***
6. Agriculture, Food and Veterinary Sciences	39—Agriculture, Food and Veterinary Sciences	0.373***
7. Earth Systems and Environmental Sciences	21—Geography and Environment	0.268
8. Chemistry	3—Chemistry	0.353*
9. Physics	9—Physics	0.446**
10. Mathematical Sciences	35—Mathematics	0.022
11. Computer Science and Informatics	22—Computer Science and Informatics	0.103
12. Engineering	12—Engineering	0.108
13. Architecture, Built Environment and Planning	37—Architecture, Built Environment and Planning	0.063
14. Geography and Environmental Studies	21—Geography and Environment	0.029
15. Archaeology	31—Archaeology	0.308
16. Economics and Econometrics	19—Economics	0.702***
17. Business and Management Studies	20—Business and Management	0.442***
18. Law	6—Law	0.306*
19. Politics and International Studies	11—Politics	-0.079
20. Social Work and Social Policy	27—Social Work and Social Policy	0.089
21. Sociology		0.274

(continued)

Table 3. (continued)

Unit of assessment	Disciplinary topic	Correlation
22. Anthropology and Development Studies	5—Sociology, Communication and Culture	0.085
23. Education	5—Sociology, Communication and Culture	0.085
24. Sport and Exercise Sciences, Leisure and Tourism	1—Education	-0.045
25. Area Studies	29—Sports and Exercise	-0.116
26. Modern Languages and Linguistics	11—Politics	0.136
27. English Language and Literature	14—Classics and Languages	0.081
28. History	26—English	-0.022
29. Classics	33—History	-0.016
30. Philosophy	14—Classics and Languages	0.496*
31. Theology and Religious Studies	2—Philosophy and Religion	0.110
32. Art and Design: History, Practice and Theory	2—Philosophy and Religion	-0.236
33. Music, Drama, Dance, Performing Arts, Film and Screen Studies	41—Art and Design	-0.153
34. Communication, Cultural and Media Studies, Library and Information Management	36—Music, Drama, Dance, Performing Arts, Film and Screen Studies	0.124
	5—Sociology, Communication and Culture	0.410**

* $P < 0.05$.** $P < 0.01$.*** $P < 0.001$.**Table 4.** A regression analysis predicting environment GPA of submissions to the Business and Management Studies panel, with the topic weightings from the eight general topics and the Business and Management topic

Predictor	Beta	R ²
Topic 4—Internal Structure of Research Units	0.013	
Topic 7—Career Development and EDI	0.010	
Topic 16—Immature Research Environment	-0.496***	
Topic 18—Staff Ways of Working	0.073	
Topic 28—REF-Focused Research Strategy	-0.083	
Topic 30—Exemplification of Strategy and Processes	0.102	
Topic 34—Industry Partners and Funding	0.336***	
Topic 40—Early Career Researcher (ECR) Development	-0.219***	
Topic 20—Business and Management	0.120*	
		0.793***

* $P < 0.05$.** $P < 0.01$.*** $P < 0.001$.

Funding (Topic 40, $\beta = 0.336$). In other words, using language associated with business and management was a stronger predictor of environment GPAs than giving examples of the unit's strategy, and around a third as strong a predictor as discussing external funding and industrial partnerships.

In sum, the more an environment statement included content from the relevant discipline, the higher the environment GPA it received, although this effect was more pronounced

for medicine, health and life sciences, and less pronounced for the arts and humanities. To illustrate this, we analysed statements with the most and the least discipline-specific language from the 'Biological Sciences' and 'Economics and Econometrics' panels (the two panels where the relationship between discipline-specific language use and environment GPA was strongest). The differing content and emphasis were clear. For example, University College London's Economics and Econometrics statement began with a sentence that listed its research strengths in "microeconomics, macroeconomics and econometrics" and went on to link their research focus to "the most pressing national and international socio-economic challenges of our time, such as inequality, migration, globalization, and sustainable growth" as well as international economic policy. By contrast, the opening remarks in the University of Northampton's Economics statement focused on being a first-time submission, and noted that this new development "has been led in part by structural changes at the University level, but more significantly by the appointment of a new Dean." Their first paragraph continued to list internal structure rather than discipline-relevant topics of strength and expertise.

Similarly, in the opening paragraph of Birkbeck's submission to the Biological Sciences UoA, discipline-specific language was used to articulate the "fundamental biological questions" conducted by staff "who are using microbial, plant and animal systems to advance our understanding of the fundamental principles underlying molecular and cellular function, physiology and behaviour" In contrast, the opening paragraphs of the University of Worcester's Biological Sciences statement described their first-time submission to the panel and articulated their internal structures and strategies (e.g. "the University went through an academic restructuring introducing Colleges and Schools"; "The University Research Strategy 2014–19 outlined the key role played by Research Groups in operationalising plans and ambitions for excellent research"). Thus, right from the start of these statements, a focus on disciplinary contribution was much clearer in the higher-scoring submissions than those with a lower GPA.

General discussion

Summary of main findings

We asked whether the perceived quality of a research environment, as measured in the UK's Research Excellence Framework, could be predicted by the text used by that unit to describe their environment. By topic modelling the full text of all 1888 unit-level environment statements submitted to REF2021, we settled on a model that included eight specific topics that were distinct from disciplinary or geographical topics. These were related to the Internal Structure of Research Units, Career Development and EDI, Immature Research Environments, Staff Ways of Working, REF-Focused Research Strategies, Exemplifications of Staff Ways of Working, Industry Partners and Funding, and ECR Development. The proportion of words each statement included from these eight topics was surprisingly predictive of the environment score that the unit received in REF2021. Specifically, these topic proportions collectively explained 58.9% of the variance in environment GPAs, and 21.9% of the variance over and above the variance explained by the unit's (standardised) FTE staff number, the (standardised)

number of doctoral degrees it awarded, and its (standardised) grant income. In total, these metrics and the topic proportions from these eight topics collectively explained 69.1% of the variance in environment GPAs. Alongside these main findings, we also identified that environment statements that contained a lot of disciplinary-specific language tended to score higher than those which did not, although this effect was stronger for medical and biological disciplines, and weaker for the arts and humanities.

Discussion

Causality

All the analyses we have reported in this paper are correlational in nature. Clearly, we were not able to experimentally manipulate the environment statements submitted to the REF and then assess the effect that these manipulations had on GPAs. Given this, care must be taken before assuming that the relationships we have reported are *causal*. Of particular concern is that some of our findings might be attributable to confounding factors. Indeed, in at least some cases this seems quite plausible. For instance, perhaps the reason why a research strategy focused on the REF seems to be negatively correlated with GPAs is that departments which have (relatively) low levels of research activity tend to both have a less mature research environment and also choose to write their environment statements using a higher proportion of REF terminology, as they have fewer pre-existing structures with pre-existing terminology to draw upon. We discuss this issue further below. Given this possibility of confounding factors, caution is required when interpreting our findings. Clearly, we cannot confidently draw causal conclusions in the absence of an experimental study (which would inevitably be of questionable external validity). Nevertheless, we can speculate.

REF environment scores are awarded through a process of human judgement. These are, by necessity given the volume of reading required of REF panellists, produced relatively rapidly. Many theories of human judgement emphasise how judgements are formed by comparing to-be-judged objects against prototypical instances sampled from memory (e.g. Fiedler 2000, 2008; Stewart, Chater and Brown, 2006; Unkelbach, Fiedler and Freytag, 2007). Such theories would likely conceptualise reaching judgements about REF environment quality as a process which involves storing multiple exemplars of high- and low-quality statements in memory and, when encountering a new statement, generating a quality estimate by matching the features of the to-be-judged statements against those exemplars or prototypes (Glöckner and Witteman 2010). This process need not be conscious, meaning that panellists are unlikely to be fully aware of the features they use to decide upon environment scores. Given this, it perhaps reasonable to suspect that if low-scoring environment statements typically have a given feature, then when panellists encounter a new statement with that feature, there may be a bias towards it receiving a lower score. In other words, if the majority of high-scoring environment statements that a panellist sees avoid using REF-heavy terminology, then in light of the decision-making literature on reasoning from prototypes and exemplars, it seems plausible that their judgements of future environment statements will be influenced by the presence or absence of this terminology, perhaps only unconsciously. If this account is correct, then the correlations between topic weightings and environment

GPAs that we have reported above may well be, in part at least, causal.

How to write a 'good' environment statement

Assuming there remains a significant narrative element to REF people, culture and environment statements in 2029, what lessons might we learn from this analysis to support the crafting of written submissions that include all those features that are associated with high-scoring statements, and none of those features associated with low-scoring statements? We make eight recommendations.

First, we would avoid stating things that high-quality research environments would consider trivial. For example, we would not mention that most staff in our unit have doctorates, or that our staff attend academic conferences and write articles in academic journals. We would avoid using the phrase "research-active", especially in an aspirational way, and we would not mention that our staff review articles for academic journals unless they had substantial editorial roles. In short, routine research activities should not be discussed in REF environment statements. Doing this is likely to give readers the impression that research is not a central feature of the unit's work and reduce perceptions of the vitality of the unit's research environment.

Second, when discussing research strategy, we would not give the impression that our research strategy is solely driven by the REF. Instead, our strategy would be organised around research centres, research groups, and departments. It would be focused on an academic discipline, not a "UoA". The staff who led our submission would not be characterised as a "UoA Working Group", and if we appointed other academic staff to REF leadership roles we would not mention it in our submission. Although we have robustly demonstrated that there is a relationship between conceptualising strategy using REF-centred terminology and receiving lower environment scores, it is less clear why this might be. One possibility is that the most research-intensive universities are sufficiently self-confident, and have a sufficiently long history of conducting research, to define their research activity in their own terms. In contrast, less research-intensive universities may need to create research infrastructure and strategies primarily in order to produce a respectable REF statement. If this were the case, we might expect the whole research enterprise in less research-intensive institutions to be more likely to be conceptualised in REF terms.

Third, we would not go into too much detail about the specific ways in which staff-related processes operate. For instance, we would not explain how decisions about sabbatical leave are informed by input from both the research committee and the teaching allocation committee. Similarly, details about which staff are involved at which stages in approving requests for conference travel funding would be omitted. Why might including detail of this sort be associated with lower GPAs? One plausible explanation is simply that including such details is a waste of space. As noted in the Introduction, REF environment statements are word limited, so including superfluous details might simply prevent the inclusion of content that would be causally associated with higher GPAs. This might be sufficient to generate a small negative relationship with GPAs (even though none would exist if there were no length restriction on submissions).

Fourth, we would not focus too much attention on how we support the career development of our ECRs. This finding is

particularly surprising in light of the REF submission guidance's statement that submissions should include "evidence of how individuals at the beginning of their research careers are being supported and integrated into the research culture of the submitting unit" (REF Panel Criteria and Working Methods 2019: 63). What might explain this apparent contradiction? An inspection of Figure 2 reveals that the negative relationship between discussing ECR development and GPA was driven by statements which included a relatively high proportion of this topic. Perhaps too much discussion of ECR development had the effect of crowding out space which could have been used for content that was more strongly associated with positive GPAs. Another possibility is that an excessive focus on ECR development might indicate to panellists that a unit feels that they have an unusually low proportion of senior established researchers in post.

Fifth, we would take care to make sure that we discussed career development and EDI, but not too much. The REF guidance emphasised that EDI should be discussed throughout submissions but, as shown in Figure 2, some submissions clearly failed to follow this guidance, and these tended to receive low GPAs. However, some submissions seemed to discuss career development and EDI too much. Our analysis suggested that devoting ~13% of the statement to this topic was optimal, with scores falling off for submissions with substantially higher or lower figures than this. Explaining why submissions which did not spend much time discussing career development and EDI tended to score poorly is straightforward: they failed to follow the clear instructions provided, and panellists may have concluded that they were poor places for minoritized colleagues to work. But why might have submissions which discussed these topics at length received lower scores? Again, one plausible account involves appealing to the length limitations of the environment statement template. Perhaps discussing career development and EDI was a qualifying criterion: not taking this issue sufficiently seriously would harm a submission, but once a submission successfully demonstrated that career development and EDI was a matter of concern, then further discussions on the topic became unnecessary. Instead, extra details on these matters had the effect of crowding out space that could have been productively used to discuss other issues associated with higher GPAs. A second possibility is that some statements mentioned EDI so often that it conveyed a 'tick box' approach rather than an authentic embedding. One final possibility is that an unusually high level of discussion of EDI issues might give the impression to reviewers that the unit felt that they had an unusually high number of issues in this area which required particular attention. Again, this might give the impression of a poor environment for minoritized colleagues.

Sixth, we would illustrate our research strategy by giving as many concrete examples as possible of how it has been implemented in practice. For example, if we provided pump-priming research funding to our staff, we would give an example of someone who had received funding, what they did with it, and what this led to. If we had particular strategies in place to facilitate interdisciplinary work, we would give an example of how this had led to a successful interdisciplinary workshop or funding application. If we had a policy on sharing research data, we might state the proportion of empirical papers in our submission where data had been shared online and give an example of how these datasets had been used by external colleagues in their own work. If we had a

particularly generous study leave allowance for colleagues returning from parental leave, we might give an example of outputs or successful grant applications that had been produced as a result of this policy, and so on.

Seventh, we would mention our research funding and industrial partnerships as much as possible. This might seem superfluous, as panellists were provided with each unit's grant expenditure alongside the written environment statement. However, our data suggest that mentioning funding and partnerships explained significant variance in GPAs over and above standardised grant income per FTE.¹ In sum, having high levels of grant income is insufficient: one must use it to provide evidence of a successful research strategy and environment as well as what the income enabled.

Eighth and finally, we would discuss our discipline as much as possible, particularly if we were writing a statement as part of a submission to a STEM UoA. For example, we would illustrate the success of our research strategy by discussing some of the important research findings it facilitated, we would name our research groups using well-understood disciplinary terms, and we would describe the work that our research funding allowed us to do, rather than merely state grant funding amounts. This finding that the use of discipline-specific language tends to be associated with higher environment GPAs is particularly interesting given the original suggestion of the Institution-Level Environment Panel Pilot that future REF exercises should abandon unit-level environment statements altogether (REF 2022). Our finding that the scores produced by discipline-specific panellists were correlated with the amount of discipline-specific language submissions used, suggests that they may have used their domain expertise to come to decisions about submission quality. Clearly this would not be possible to the same extent, if at all, if environment were assessed at the institutional level by a panel made up of experts from a variety of disciplines—even if, as the REF (2022) pilot panel proposed, brief unit-level narratives were incorporated into the institutional-level statement. In sum, our results indicate that research environment assessed at the institutional level would likely be a different construct from research environment assessed at the unit level.

It has now been communicated that the next Research Excellence Framework will continue to assess an institution's people, culture and environment at both institution- and discipline level, with the greater weight being given to the discipline-level assessment (Joint UK HE Funding Bodies 2023). It is not yet clear the extent to which this element will look beyond the domains assessed in REF2021 (context, people, income & infrastructure, and collaboration & contribution), nor the extent to which the assessment will consist of quantitative indicators relative to narrative description. Regardless, our analysis may help institutions reflect upon what it means to have a high-quality research environment.

Acknowledgements

We are grateful to Hugues Lortie-Forgues, Victoria Simms, Steve Rothberg, and two anonymous reviewers for insightful comments on earlier drafts of this manuscript.

Funding

This work was partially supported by Research England, via an Expanding Excellence in England grant to the Centre for

- Thomas, D. A., Nedeva, M., Tirado, M. M., and Jacob, M. (2020) 'Changing Research on Research Evaluation: A Critical Literature Review to Revisit the Agenda', *Research Evaluation*, 29: 275–88.
- Thorpe, A., Craig, R., Tourish, D., Hadikin, G., and Baticic, S. (2018a) 'Environment' Submissions in the UK's Research Excellence Framework 2014', *British Journal of Management*, 29: 571–87.
- Thorpe, A., Craig, R., Hadikin, G., and Baticic, S. (2018b) 'Semantic Tone of Research 'Environment'submissions in the UK's Research Evaluation Framework 2014', *Research Evaluation*, 27: 53–62.
- Unkelbach, C., Fiedler, K., and Freytag, P. (2007) 'Information Repetition in Evaluative Judgements: Easy to Monitor, Hard to Control', *Organizational Behavior and Human Decision Processes*, 103: 37–52.
- Watermeyer, R., and Derrick, G. E. (2022) 'Affective Auditing: The Emotional Weight of the Research Excellence Framework on Middle Management', *Research Evaluation*, 31: [rvac041].498–506. <https://doi.org/10.1093/reseval/rvac041>