July 2022

REF Institutional-level environment pilot: Report of the pilot panel

REF 2021 Research Excellence Framework

Contents

Executive summary

1. Alongside the UK's Research Excellence Framework (REF) 2021, the four UK higher education funding bodies ran a pilot exercise to explore the feasibility, benefits and drawbacks of an assessment of institutional-level (IL) environment, to inform whether and how to include this element in future exercises. The pilot was introduced in response to a key recommendation of the independent review of REF 2014, led by Lord Stern in 2016, which aimed to reduce duplication across multiple unit-level submissions in REF and enable the process to capture aspects of the institution's environment that reflect institutional-level activity.

2. This report sets out the findings of the institutional-level environment pilot panel (ILEPP). It sets out our conclusions and recommendations to the funding bodies, to inform decisions about IL assessment in future research assessment exercises.

3. The pilot assessment took place alongside the main REF 2021 exercise and did not contribute to the outcomes of that exercise. It included assessment of the narrative IL environment statement (REF5a) and supporting environment data provided by institutions participating in REF 2021. The IL statement provided institutional context for the unit-level (UL) environment assessment undertaken by each sub-panel but was not itself scored as part of REF 2021.

4. Having undertaken the process of developing the guidance and criteria, assessing the REF5a IL statements, and producing and analysing the assessment outcomes, our key conclusions are that:

    a. We were able to apply the assessment criteria, differentiate between submissions and produce a quality profile for each. Therefore, the pilot process has demonstrated that the extension of the environment element of REF to assessment at the level of the whole institution is feasible. Furthermore, in the context of the purposes of the REF, there is clear value to be derived from assessment at this level.
    b. Assessment at the institutional level is a more appropriate model for the research environment in the REF than assessment at the level of the submitting unit.
    c. Several refinements will need to be made to the submission and assessment processes in future at this level, in place of a unit-level assessment. These would be needed to ensure the

process is sufficiently robust and equitable for contributing to REF outcomes and, consequently, informing the allocation of funding for research.

5. We make a number of recommendations for including and refining IL environment assessment in future exercises. These are summarised below and explained in more detail throughout the report.

Recommendations

Overall

R1. For future research assessment, the environment element should be assessed through a single institutional-level submission. This reflects the clear opportunities for reducing burden for submitting institutions and the high value offered in requiring information about the institution as a whole. It is further supported by the broad alignment observed between the outcomes from the pilot and the average UL environment outcomes. The detailed submission requirements should be refined in consultation with the sector.

R2. The submission should incorporate and enable understanding of contributions and brief key information from each of the submitting units within the HEI.

Submissions

R3. All participating institutions should follow a single format for IL environment submissions.

R4. Work with sector representation should seek to identify the impact of any changes for smaller/less-research-intensive HEIs and consider how these could be addressed.

R5. The submission template and guidance should be more tightly specified, including standardised supporting data. This approach should also be extended to the UL information to be provided as part of the IL submission.

R6. All elements of the submission should be scored; the context section in the current template could move into the strategy section.

R7. Detailed requirements should be developed in consultation with the sector, with an aim to provide initial guidance at an early stage.

Assessment

R8. The starred level definitions should be reviewed to support a broad definition of excellence and to consider the relationship between supporting research and enabling impact.

R9. Calibration, moderation, and final review assessment processes should be extended in a future assessment.

Panel

R10. A single IL panel should be appointed to undertake assessment across all submitting institutions.

R11. The IL panel should be constituted to provide expertise at the appropriate level to assess the wider institutional environment, including higher managerial, strategic and administrative, research user, and small and specialist experience.

R12. The panel should be extended to include specific expertise in research integrity issues and EDI legislation and practice.

R13. The assessing panel should work with the sub-panels for advice on disciplinary-relevant elements of the submissions.

REF Institutional-level environment pilot: Report of the pilot panel

Foreword

During my career in UK higher education and research I have overseen three submissions into REF and, before that, RAEs, concluding with the most recent as Vice-Chancellor of Newcastle University. I have also served on the REF panels, acting as chair for the Clinical Medicine sub-panel in REF 2014. Throughout this time, I have seen first-hand the amount of work involved in preparing – and assessing – environment templates for each submitting unit returned. I was attracted, therefore, to the initial proposal set out in the Stern review for an institutional-level assessment of the research environment. Should it prove feasible, it seemed to me an opportunity to get a view of what was happening across the whole institution to support and develop research, without the need for all the words generated for the multiple unit-level assessment.

I was therefore delighted to be invited to chair the panel for the pilot assessment. The process ran very smoothly, with excellent contributions and insights provided by the pilot panel, who among them have a considerable range of expertise in managing and supporting research and enabling its impact. The contributions from our international members were also incredibly valuable to our process. I extend my thanks to all the members for the time and effort they have given to this exercise.

As is the nature of a pilot exercise, we began this process unsure of what we would find. The outcomes have been both pleasing and surprising in equal measure, showing that, with some tweaks, the institutional-level environment has much to offer future research assessments. As envisioned by Stern, saving burden is a key offering; on top of that, our panel saw the potential for an assessment that recognises and rewards the establishment and maintenance of a strong institutional level research and impact environment in an approach that's more closely aligned with the way that the consequent funding is distributed. We hope our findings and conclusions are helpful for the funding bodies as they begin to set the framework in place for future assessment.

Professor Chris Day FMedSci
Chair of ILEPP and Vice-Chancellor and President, Newcastle University.

Introduction

1. A key recommendation of the independent review of REF 2014, led by Lord Stern in 2016[1], was the introduction of an institutional-level (IL) submission describing the institution's strategy and support for the research environment. This recommendation aimed to reduce duplication across unit-level submissions, enable the accurate representation of aspects of an institution's environment that reflect institutional-level activity, and capture institution-wide strategic objectives and cross-cutting structures and initiatives.

2. Following initial REF consultation in 2017, the four UK higher education funding bodies set out their decision to formalise the inclusion of IL information at the unit level, and to pilot the standalone assessment of the IL environment alongside the REF 2021 exercise. The submission of an IL environment statement (REF5a) was a requirement for all institutions participating in the REF, with an exception for those submitting in one unit of assessment (UOA) only, where submission of the REF5a was optional. The IL statement provided institutional context for the unit-level environment assessment undertaken by each sub-panel but was not itself scored as part of REF 2021. The pilot was undertaken separately by a panel constituted for the purpose and did not contribute to the REF outcomes.

3. The institutional-level environment pilot panel (ILEPP) was established in November 2018 to conduct the pilot of the standalone assessment of IL environment submissions and to advise the REF team and the funding bodies on the feasibility, benefits and drawbacks of IL assessment, to inform whether and how to include this element in future exercises. The pilot assessment was conducted between June 2021 to May 2022, and included a calibration stage, assessment of submissions and moderation, and a review and recommendations stage. ILEPP was chaired by Professor Chris Day, Vice Chancellor of Newcastle University. Details of the panel membership are set out at Annex A.

4. This report describes ILEPP's findings on the pilot assessment, and presents the panel's conclusions resulting from this. In the final section, the report makes a number of recommendations based on the pilot assessment, to inform decisions about IL assessment in future exercises. As agreed through consultation with the sector prior to the start of the pilot, quality profiles for individual institutions will not be published as

[1] https://www.gov.uk/government/publications/research-excellence-framework-review

part of the pilot outcomes[2]. The panel will provide individual feedback confidentially to the heads of all submitting institutions. This does not include the quality profile but provides a narrative commentary on the institution's submission.

[2] 'Institutional-level environment pilot: supplementary guidance on submissions and panel criteria and workingethods' (REF 2019/06). Available at www.ref.ac.uk under 'Publications and reports'.

Section 1: Submissions and assessment

Pilot submissions

5. Each institution participating in REF 2021 was required to submit a narrative statement (REF5a) describing the institution's strategy and resources to support research and enable impact, relating to the period 1 August 2013 to 31 July 2020. Small and specialist institutions that made a submission in one UOA only were not required to provide a REF5a statement but could choose to submit one where this was the most appropriate way of representing the institution's research environment. Seven institutions opted not to submit the REF5a.

6. The REF guidance and criteria described the requirements and provided the submission template for the IL environment statement (REF5a)[3]. Supplementary guidance was set out by the pilot panel in 2019, which was intended to support institutions' development of the IL environment statement, and set out details of information and indicators for inclusion.

7. Each submitted REF5a statement consisted of four sections, covering the following: context and mission; strategy; people; income, infrastructure and facilities. In providing evidence in the REF5a statement, institutions were advised to draw on supporting quantitative indicators where applicable, and were encouraged to refer to the advice and examples of indicators provided by the Forum for Responsible Research Metrics (FFRRM)[4]. The total length of the statement was determined by the number of full-time equivalent (FTE) staff returned by the institution across all its submissions, ranging from 4,000 words for the institutions returning under 100 staff (FTE) to 5,500 for those returning 1,000 FTE or more

8. As part of the main REF exercise, for each submission it made in a UOA, an institution submitted environment data relating to the assessment period, including the number of research doctoral degrees awarded (REF4a), external research income (REF4b) and research income-in-kind (REF4c). For the purposes of the pilot, these data were aggregated across an institution's submissions and accompanied the REF5a statement,

[3] The 'Guidance on submissions' (REF2019/01) and 'Panel criteria and working methods' (REF2019/02) are available at www.ref.ac.uk under 'Publications and reports'.
[4] The FFRRM guidance is available on the REF website: https://www.ref.ac.uk/guidance-and-criteria-on-submissions/guidance/additional-guidance/

along with further data aggregated from the standard analysis described in Annex J of the 'Guidance on submissions' (REF2019/01).

9. In 2020 additional measures were introduced into REF to take account of the effects of Covid-19 on submission preparation. Institutions were invited to provide an annex to the REF5a statement (max. 500 words), describing the particular changes affecting their environment as a result of Covid-19 and how the institution has responded, in the final part of the assessment period.

Pilot criteria and working methods

10. The assessment undertaken by the panel was criteria based, reviewing each submission against the stated criteria of 'vitality' and 'sustainability', as applied in the main REF exercise. The panel undertook to judge each submission on its merits, contextualised appropriately to the nature of the institution. The panel's assessment did not use or refer to subgroupings of institution.

11. In building up a graduated quality profile for each submission made, ILEPP assessed the information provided in the REF5a statement, or REF5b unit-level template, where no REF5a was submitted, using the starred quality levels identified in the 'Guidance on submissions', Annex A. The panel considered the environment data and Covid-19 annex (where submitted) within the context of that information. In common with the sub-panels undertaking assessment of the environment, a section of the statement that was judged to be on the midpoint between two of the starred levels was assigned a half-point grade (e.g. 3.5, 2.5, 1.5 or 0.5). Where this occurred, that section of the environment template contributed to the environment sub-profile by assigning half of its grade to each of the two starred quality levels that the midpoint grade fell between.

12. As set out in ILEPP's supplementary guidance and criteria, section one of the REF5a: Context and mission would not be scored and would provide background information to support contextual assessment of the other sections. For each of the remaining three sections an equal weighting would be applied. For the seven submissions where the REF5b was assessed, the relevant parts of that template were considered against the IL criteria, and a score was recorded against each of the three IL sections.

13. ILEPP undertook its assessment in accordance with REF procedural guidance for managing conflicts of interest. No panel member was involved in review of any submission for which they had declared a major interest. In all panel discussions pertaining to assessment of particular

submissions, conflicted members were required to be absent from the meeting and this was recorded by the panel secretary.

14. In line with the practice adopted by the main and sub-panels, ILEPP members undertook tailored 'Fairness in REF' training prior to commencing the assessment process, and developed an intention plan to mitigate against bias in the process.

15. An initial calibration exercise was undertaken across the panel, using a sample of submissions with which the panel collectively had no major conflicts of interest. This aimed to develop a shared understanding of the starred quality levels and application of the criteria. Following completion of the calibration, the chair, working with the panel secretariat, allocated submissions for assessment. Submissions were allocated across five sub-groups, ensuring a mix of institutional type, size and specialism, a balance of institutions across the four UK nations, and taking into account conflicts of interest. The international members and the chair undertook a moderation role, reading across groups and during the moderation exercise described below.

16. Assessment was undertaken by members individually, before meeting together in sub-groups to review and agree scores and feedback for each allocated submission. Consistency of assessment standards across the groups was examined through detailed analysis of scoring, and through a moderation exercise early on in the assessment process. In this exercise, a sample of submissions was assessed by two other sub-groups, and/or international members and the chair. The quality profiles and feedback for each submission were reviewed and agreed collectively by the whole panel, taking account of conflicts of interest.

17. During the assessment process, ILEPP received advice from the REF Equality and Diversity Advisory Panel (EDAP) on the overall strengths and weaknesses observed in relation to equality, diversity and inclusion in EDAP's review of the IL environment statements. The advice informed ILEPP's assessment of the 'People' section of the submissions.

18. Following completion of the assessment, ILEPP undertook analyses of outcomes to identify trends and patterns in the results, and to examine the alignment of the IL pilot outcomes with the published unit-level quality sub-profiles for the environment, to inform the panel's conclusions and recommendations.

Section 2: Assessment outcomes and observations

Quality profiles

19. Across 157 submitting institutions, we assessed the quality of the IL environment for over 76,000 academic staff (FTE) employed with a significant responsibility for research. This ranged from submissions with just under 4 staff to the largest at over 3,400. For each submission we produced a quality profile based on our scoring of the three equally-weighted sections of the template, informed by the standard analysis and Covid-19 annex as relevant.

20. We judged that 38.3 per cent of UK university research environment activity was conducive to producing world-leading research and enabling outstanding impact. The average FTE-weighted quality profile for the IL environment is shown at Table 1[5].

Table 1: IL environment average weighted profile

FTE-weighted average percentage of all submissions judged to meet the standard for:				
4*	**3***	**2***	**1***	**U/C**
38.3	41.1	16.8	3.6	0.2

21. Further analyses of the data show that excellence was identified in institutions of all sizes, with examples of the highest scores found across small to very large institutions. We did, however, observe some relationship with size and quality, as shown by quartile data in terms of the proportion of 4 star: the top quarter of institutions had an average of almost 1,200 FTE, compared with just 40 FTE for the lowest quarter.

22. Analysis of the outcomes shows a relationship between the quality of the research environment as assessed through the REF5a and the research intensity of the institution. For the purposes of analysis of the outcomes,

[5] The average profile is weighted by staff FTE to take into account the relative size of submissions in the overall outcome. It is produced by weighting the proportion of activity at each starred level for each submission by the total FTE of Category A submitted staff returned by the institution.

a broad indicator of research intensity provided by the TRAC peer groupings was used. We observed that the highest weighted average proportion of activity judged at 3 star or above achieved by institutions in Peer Group A (institutions with a medical school and research income of 20% or more of total income), with Peer Group B (all other institutions with research income of 15% or more of total income) achieving the next highest[6].

23. We observed a disparity in the level of detail that could be provided by small and specialist institutions returning a REF5b, which has a much greater word limit than the REF5a, particularly in contrast with those small and specialists who returned a REF5a. In these latter cases, frequently the extent of cross-referencing to detail in the REF5b significantly limited our ability to score the REF5a content. In terms of outcomes, those returning only the REF5b tended to achieve higher scores, with the weighted average proportion of 4 star for these HEIs over 7 percentage points higher than the weighted average for the single-UOA group that provided REF5as.

Features of institutional-level submissions

24. Across the set of submissions we reviewed, we observed wide-ranging evidence of environments that were supporting world-leading research in range of settings. Each submission was judged on its merits, referencing the nature of the institution in respect of size, mission and focus. We based our assessment solely on the submitted material provided and observed a significant variation across the submissions in terms of the quality of presentation and evidence.

25. Inevitably with the narrative basis of the submission, the strength of the evidence presented plays a part in the assessment. The best examples were able to explain and evidence claims in clear terms, covering relevant aspects of the institutional environment. For others this was less evident, relying on extracted details from pre-existing documents and strategies, failing to provide supporting evidence and at times reflecting a less concentrated focus on producing the IL submission. Some exhibited different writing styles across the document, suggesting limited central oversight of the submission in several cases. We have sought to summarise the characteristics we observed across the stronger and weaker cases further below.

26. One factor recognised by the panel throughout its assessment was the effect of Covid-19 on HEI preparations for submission. This may have

[6] Research income is defined as the funding council recurrent research grant plus the total research grants and contracts returned in the HESA Finance Statistics Return (FSR), 2012-13.

impacted negatively on some institutions' preparations. We recognised that while REF5a was a required element, it was not scored in the formal assessment; it was provided to sub-panels, to inform and contextualise the assessment of the UL template. Understandably, this may have led to its deprioritisation in some instances, in order to focus resources more on the scored elements.

Small and specialist submissions

27. As noted above, one of the key observations we made during the process was around the issues encountered in assessing several of the submissions by small and specialist institutions, who returned in one UOA only. This was manifest in two ways:

    a. REF5b: we reviewed seven REF5bs as part of the assessment process from single-UOA institutions that had chosen not to submit a REF5a. The word limits for this template were different to those for REF5a, in effect allowing a sub-set of the smaller institutions significantly greater space than all others and therefore giving a greater level of detail in the evidence provided. However, as the focus of the REF5b was towards the formal element of the UL assessment, it was not well suited to assessment at IL requiring extrapolation of relevant information by the ILEPP assessors.

    b. REF5a: for small and specialist institutions that chose to return a REF5a, we observed that in some cases significant detail relevant to our assessment had been set out in the REF5b and cross-referenced to from the REF5a. As the focus of our assessment was on the material submitted in the IL process, the limited information available to us in these cases affected the grading.

Quantitative indicators

28. In ILEPP's supplementary guidance for institutions published during the submissions phase of REF in 2019, institutions were requested to provide evidence and indicators relevant to their own context, and encouraged to consider the guidance from the FFRRM in so doing. We noted that we would expect to receive three key indicators: recruitment by age; professors and senior staff by protected characteristic; and gender pay gap.

29. The panel noted from the submissions that the response to this requirement was mixed, with significant variation in the quality and scope of data provided. There were many examples of strong use of data, appropriately contextualised and linked to the narrative commentary – particularly around achievements and outcomes. However, we also noted

that for a significant proportion of submissions supporting data was infrequently used, or less complete in key areas, for example with limited information in many submissions on protected groups beyond gender. The inconsistency in the use of quantitative evidence increased the complexity of and potential subjectivity in the assessment process.

30. As part of the assessment process, the panel was provided with the aggregated data in the standard analysis, including the environment data (REF4a/b/c) that were submitted as part of the main assessment. While the provision of data in a standardised format was helpful, we observed some disparity between the submitted metrics and the information set out in the statement. The presentation of the income data as an annual average, due to a financial reporting change in 2015, also made it difficult to discern trends across the REF assessment period.

Characteristics of submissions

31. Some of the strongest submissions displayed some or all of the following:

    a. **Clarity, coherence and appropriateness of strategy:** strong submissions articulated a clear strategic direction, relevant and appropriate to the overall focus, size and specialism(s) of the institution, with coherent plans towards their achievement. Examples were seen across those with a focus more aimed towards the HEI's local area or wider region, and those with a more international outlook in terms of partnerships and activities.

    b. **Self-reflection on EDI issues:** institutions able to demonstrate a level of critical reflection on their approach to wider staffing, across all levels, and in particular on their progress and plans around equality, diversity and inclusion tended to score well against the criteria. Strong examples also tended to go beyond gender alone in considering EDI issues. As with other aspects, it was important that claims around progress achieved were sufficiently and appropriately evidenced.

    c. **Reporting outcomes with appropriate evidence:** a key strength noted in some submissions was a focus on reporting outcomes achieved. The focus of these outcomes may be within one or more of the institution's units, or may reflect wider outcomes across the HEI.

    d. **Achievement of objectives clearly set out and evidenced:** we noted that strong submissions articulated achievement of strategic objectives for their institution, contextualised within the overall strategic direction for that institution.

e. **Effective use of indicators, examples and evidence:** the narrative-based assessment approach relies centrally on the use of evidence and examples to support the claims being made within the statement. Strong submissions ensured this was done effectively, drawing on appropriate indicators, throughout the statement. Strong submissions often showed clear structure and presentation, with effective use of sub-headings and word counts that were balanced across the scored sections.

32. Some of the weaknesses identified across the submissions included:

f. **Over-detailed context and mission statements:** many submissions – reflecting a range across the quality levels – included key information and/or too much detail in the 'Context and mission' section, which is not a scored section and limited the word count available for the remaining, scored sections.

g. **Strategies, policies or activities limited in scope:** we noted in several cases that aspects described in the IL environment did not demonstrate ambition or achievements in line with the wider context of the institution.

h. **Lack of clarity around how content contributes to or supports research:** we noted examples where there was a lack of clarity around the way in which activities, facilities or achievements described in submissions contributed to supporting research and enabling its impact. For example, highlighting investments in facilities which related to supporting teaching activities rather than research.

i. **Lack of clear structure, focus and / or detail:** for some submissions the presentation of the material significantly affected our ability to assess the content of the submission – including in the way the evidence was structured, the amount of detail provided and the extent to which the narrative provided a clear focus. In some cases hyperlinks were provided in lieu of explanation, which could only be followed by the panel in the event of audit.

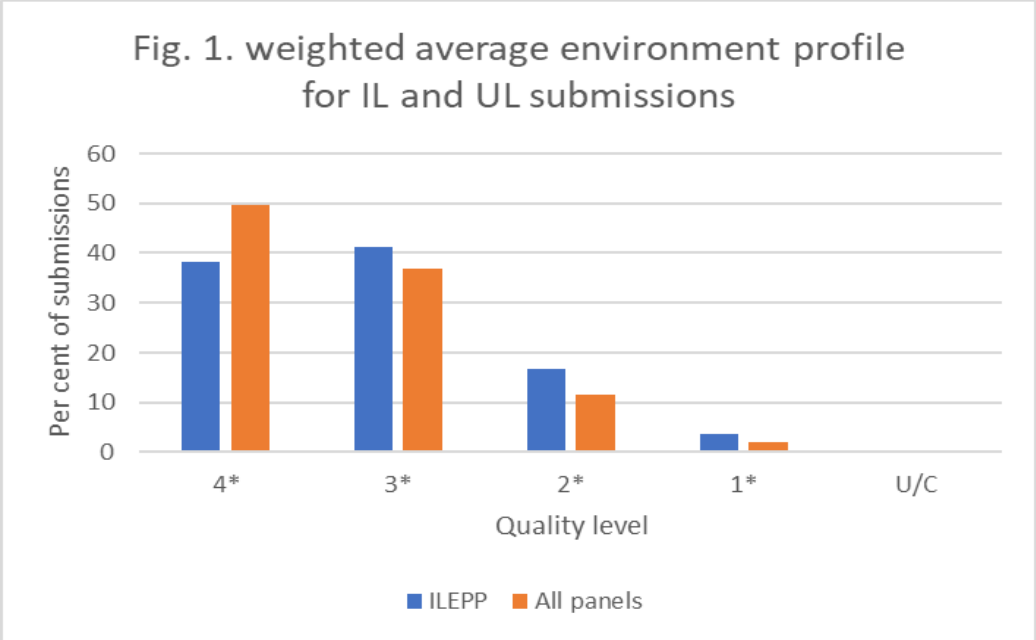
j. **Variable detail / evidence across different sections:** the assessment of some sections of the statements was limited by inconsistent levels of detail and evidencing provided, so that in some instances there was very little material to review. This was particularly evident for the last section on income, infrastructure and facilities.

k. **Reliance on description, rather than on outcomes and evidence of achievements:** some weaker examples provided primarily descriptive accounts of activities and aims, which while helpful for understanding context, did not sufficiently articulate what had been achieved. In some cases, there was a general lack of, or unclear use of, evidence and data to support claims. Several statements included hyperbolic statements about quality and excellence, which were themselves not substantiated by use of evidence and examples.

Reviewing IL and unit-level submissions

33. When the UL environment sub-profiles were available in May 2022, we undertook analysis to compare the outcomes as part of the pilot process. This was undertaken for two key reasons: firstly, to understand the degree to which the single IL quality profile for an institution corresponded with the weighted average UL environment sub-profile for an institution; and secondly, to understand the extent of variation between an institution's UL sub-profiles, and the relationship of this with the single IL profile.

34. Figure 1 shows the weighted average ILEPP profile compared with the weighted average UL environment sub-profile for all submissions in REF 2021.



35. As the chart shows, while the distribution of quality across the starred levels is broadly consistent, the data highlight the lower average scores emerging from the pilot than those awarded in the main assessment. In

particular, there is an 11 percentage point difference in the proportion of four star observed between the ILEPP and all panels weighted UL environment profiles. There are likely to be a range of reasons for this, including:

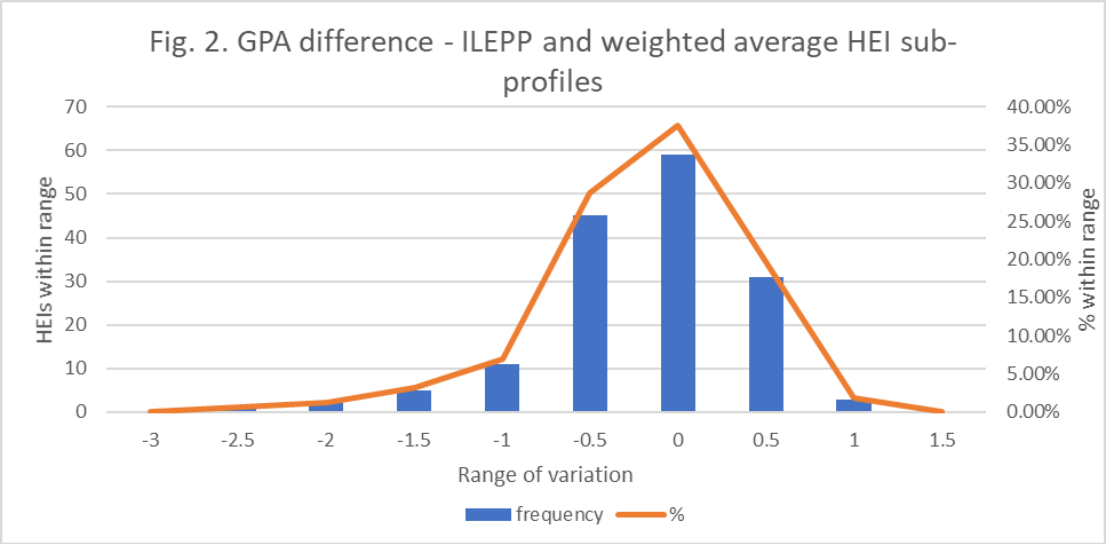
a. The REF5b UL environment template reviewed by the sub-panels was assessed on four sections – including 'Collaboration and contribution to the research base, economy and society' as a further section not covered in the IL statement. This may have provided additional opportunity for submissions to achieve higher scores, as well as more generally reducing the weight each section contributes to the overall profile in comparison to the three scored sections in the IL assessment. There were also differences in the weighting of the four UL sections in Main Panel D, with the 'People' section weighted at 30 per cent, while 'Income, infrastructure and facilities' was weighted at 20 per cent.
b. While the submitted REF5as were provided to the assessment panels in the main exercise, these statements were not scored but were used to inform and contextualise the assessment of the REF5b. We noted considerable variability in the quality of submissions, and observed that in some cases comparable investment had not been made by HEIs in preparing the REF5a as with other aspects of submissions. This meant that in some submissions presentation, content and supporting data were poor, which negatively impacted on scores awarded.
c. Relatedly, we anticipate that the timing of the emergence of Covid-19 may have negatively impacted on some institutions' preparations, including the types of moderation and review processes involved in drafting institutional documentation.
d. A number of submissions, including some of better quality, included key information in the unscored 'Context and mission' section of REF5a, which also negatively impacted on scoring.

36. To examine the variation between IL profiles and the aggregated sub-profiles for submitting institutions, a single data point was required for each profile. For this purpose, we primarily used the grade-point-average (GPA) to undertake this analysis[7]. However, it should be noted that as with any summary measure, the GPA will reduce the richness of information that the quality profile provides. It is often useful, therefore,

[7] The GPA is calculated by multiplying the percentage of activity judged at four star by 4, three star by 3 and so on, then summing these together and dividing by 100 to produce a number between 0 and 4.

to use additional measures that can provide a further insight – such as the percentage of activity judged at 3 star or above.

37. Figure 2 plots the frequency of institutions by the difference between the institution's ILEPP GPA and the weighted average GPA of environment sub-profiles awarded across that institution's submitting units.



38. The majority (85 per cent) of the GPAs for the IL profiles fall within a ±0.5 difference with the HEI's weighted average GPA for the UL environment profiles. This skews more towards the negative – with ILEPP scores more frequently showing a lower GPA – which is consistent with the lower average scores emerging from the pilot assessment, as observed above. The institutions achieving a higher ILEPP GPA include a mix of institution sizes and intensity, although as reviewed above, this group was more likely to include larger, research intensive institutions. To some degree, this reflects the different ability to achieve full or very near full marks on the assessment of a single template with three scored sections in comparison to an average across a large number of submissions. Within the wider group of institutions falling ±0.5 away from the average HEI GPA, there is no notable relationship between the extent of variation and the quality level.

39. Those institutions at the extreme ends on the variation scale, including those where the ILEPP GPA was either 1.5 or more beneath the weighted average HEI GPA, or at least 1 above, are almost entirely small and specialist institutions. This is likely to be reflective of some of the issues encountered during the assessment process for this type of institution, where the submission of a REF5a was optional, as we described above.

40. We also examined the small group of institutions with an ILEPP GPA of between 1 and 1.49 beneath the weighted average HEI GPA. Here in

particular we noted a correspondence between the extent of variation and the limited amount of evidence presented in the REF5a statement, reflecting comparatively lower levels of investment in its completion for the purposes of the pilot.

41. The remaining analysis was therefore focused on investigating in more depth the characteristics of institutions with greater or less variation among the ±0.5 group. The average difference in the ILEPP GPA from the weighted average HEI GPA was greatest for institutions in middle two quartiles, according to total submitted FTE. This also showed a lower average difference for the largest institutions, and to a smaller extent, for those in the first quartile. To further enhance our understanding of the outcomes, analysis was also undertaken on the percentage point difference in the proportion of the ILEPP profile judged at three star or above compared with that of the weighted average HEI sub-profile. This showed a similar trend to the GPA variation. These data are summarised in Table 2.

Table 2: Difference in IL and average weighted UL outcomes (GPA and proportion of 3* and above), by HEI size

Quartile (by total staff FTE)	Average difference in GPA between ILEPP and weighted average HEI profiles	Average percentage point difference in proportion of 3* and above
Q1	0.38	14.3
Q2	0.41	24.6
Q3	0.43	25.0
Q4	0.24	6.4
Total (n=135)	**0.36**	**17.5**

42. Linked to submission size, we also examined variation in terms of research intensity. While those institutions in TRAC Peer Group A showed the lowest average variation (GPA and three star or above), institutions in groups D and B showed the highest average variation.

43. Noting feedback from the REF panels about the relative under-representation of the arts and humanities in the REF5a statements, analysis was conducted to look at the relationship between the proportion of an institution's submissions into Main Panel D UOAs and the extent of variation between the ILEPP profile and the weighted average HEI profile. When factors relating to HEI size and research intensity were accounted for, the data did not show a notable relationship between the degree of variation and a higher proportion of

submissions in Main Panel D. However, this analysis did identify a group of HEIs whose research environment is currently emerging, whose discipline focus tends more towards Main Panel C and D disciplines, and whose REF5a statements showed several of the limitations described above – including limited focus on outcomes, evidence and insufficient detail.

Variation within an HEI's average profile

44. The average weighted profile for an HEI provides some insight into the overall proportion of staff with significant responsibility for research are working in environments of a given quality – for example, an average weighted profile showing 60 per cent at three star, 20 at four and 20 at two star, would indicate that the majority of staff in the institution are working in an environment primarily conducive to producing research of internationally excellent quality and enabling very considerable impact.

45. Analysis also shows that there is considerable variation across the individual UL environment sub-profiles, and that this variation is greater for some sizes and types of HEI than others. Table 3 shows the average range between the minimum and maximum GPA of an institution's environment sub-profiles, by institution size. This excludes institutions who submit in one UOA only, and hence do not have multiple profiles across which variation can arise.

Table 3: Average range in GPA of UL environment submissions, by HEI size

Quartile (by total staff FTE)	Average of GPA range across an HEI's UL profiles
Q1	1.18
Q2	1.56
Q3	1.67
Q4	1.31
Grand Total	**1.47**

46. The data show that institutions in the middle two quartiles by size (ranging from around 80 FTE to around 630 FTE) have the greatest range in GPA. This is further underlined when looking at range by number of submissions, with HEIs submitting in 11 to 20 UOAs showing the highest average range in GPA. When considering by research intensity, Peer

Groups C and D show the greatest average ranges in GPA between submissions[8].

47. While the higher ranges for Q2 and Q3 may indicate some relationship between the range of quality across UL profiles within an HEI and the degree of variation observed between the ILEPP GPA, this is not borne out by more detailed examination of the correlation between these figures.

48. For institutions submitting in two or more UOAs, the ILEPP GPA shows an average difference of 0.38 from the mean GPA of the UL submissions. In line with the wider scoring trend observed in the overall profile, the ILEPP GPA is more likely to be closer to the minimum GPA of an institution's UL sub-profiles than the maximum; although for most of these cases, the ILEPP GPA is the higher value. It is also notable that the ILEPP GPA showed the highest maximum score (4) in a small number of cases, which was not observed in the mean (or weighted) GPA for HEIs submitting in more than one UOA.

[8] Peer group C: Institutions with a research income of between 5% and 15% of total income; Peer group D: Institutions with a research income* less than 5% of total income and total income greater than £150M.

Section 3: ILEPP's Reflections and recommendations

49. Taken as a whole, the panel considered that the submission requirements, assessment approach and panel composition were appropriate for the purposes of the pilot exercise, allowing us to undertake and evaluate the process for future implementation. Conducting the pilot also allowed us to identify where changes could be made in any future assessment process. This section sets out our reflections on the process and our formal recommendations to the funding bodies on the inclusion of IL environment submissions in future assessment exercises.

Overall conclusions

50. Having undertaken the process of developing the guidance and criteria, assessing the REF5a IL statements, and producing and analysing the assessment outcomes, we have concluded that:

    a. We were able to apply the assessment criteria, differentiate between submissions and produce a quality profile for each. Therefore, the pilot process has demonstrated that the extension of the environment element of REF to assessment at the level of the whole institution is feasible. Furthermore, in the context of the purposes of the REF, there is clear value to be derived from assessment at this level.
    b. Assessment at the institutional level is a more appropriate model for the research environment in the REF than assessment at the level of the submitting unit.
    c. Several refinements will need to be made to the submission and assessment processes in future at this level, in place of a unit-level assessment. These would be needed to ensure the process is sufficiently robust and equitable for contributing to REF outcomes and, consequently, informing the allocation of funding for research.

The value and opportunity of a single IL environment assessment

51. Where constructed effectively, the IL submissions provided a unique view of the institutional activities, strategies, objectives and achievements of the research environment for the submitting institution. As intended by Stern, this allowed recognition of the investment in and management of research at the institution level.

52. We therefore considered that the assessment at the IL provides a more rounded understanding of the overall environment, strategic direction and across-the-piece support for staff, use of resources, planning and engagement of an institution than is provided by unit-level information – either in isolation or reviewed collectively. This appropriately reflects the responsibility that lies with the institution for creating and maintaining the structures of a supportive research environment. In the context of the resource allocation and accountability purposes of REF, with funding allocations made on an unhypothecated basis to each submitting institution, we judged that the provision of detailed information at this level was of high value.

53. In section 2, we reported our analysis of the pilot IL outcomes and the UL outcomes aggregated across each submitting HEI in REF 2021. This highlighted that the ILEPP GPA for the majority (85 per cent) of HEIs was within ±0.5 points of the average weighted GPA for the UL environment. This was negatively-skewed, with ILEPP outcomes more likely to be lower than the average weighted institution profiles. Some of the more significant differences observed have likely explanations in the context of the pilot – including the process issues for the small and specialists HEIs and the challenges encountered across a wider set of submissions around sufficient evidence and detail. The majority of institutions fell within a narrower range of difference in GPA, with increased variance more likely for institutions that sit in the middle two quartiles in terms of size.

54. It is clear that a single profile produced through an IL assessment will not be able to replicate the potential diversity of outcomes achieved by an HEI returning in two or more UOAs. The outputs and impact profiles would continue to maintain a UL view of quality; further analysis could seek to examine the difference in the overall quality profiles at UL with and without the environment element. The refinements we propose in this report may also address some of the negative differences observed between the ILEPP and average outcomes. But it is our view that the IL assessment should not in any case be seeking to replicate the UL process.

55. Unit-level assessment allows recognition of excellence at a more localised, disciplinary level within the context of the wider discipline, whereas IL assessment focuses more towards recognising excellence in the overall strategic management of research, staff support and investment across the HEI. While in many instances these factors align to highlight excellence at both levels, this will not always be the case where local practice, whether viewed positively or negatively, is not replicated or consistent at IL.

56. We recognised the value of the detailed information on the research environment provided in the current model, and that in many cases this information may be drawn on to drive internal funding allocations within the institution. We considered it would be important to retain some of the nuance and granularity provided through the UL statements in a single IL environment model through the inclusion of key data and brief narrative at UL. In this approach, specific attributes of individual unit environments can continue to be highlighted. We also considered the richness of the information provided by the assessment of output and impact at UL, which would continue to allow allocation of funding internally.

57. The panel also reflected on the potential benefits arising from the submission process in having an element of the assessment where the HEI centrally communicates its strategic approach. This moves away from potential inconsistencies or an incomplete picture across individual submissions, and may also remove less relevant information, allowing for greater concision and more focussed use of resources. ILEPP considers that an effective submission would require central ownership of the IL environment submission, coordinating activities and engaging with its submitting units in its preparation.

58. There is a wider burden-saving aspect to the preparation of a single IL submission for the environment, in contrast with the preparation of multiple UL templates across an institution's submitting units. ILEPP members involved in or having oversight of UL submissions within their own institutions, reflected on the resource commitment that this represented, with multiple iterations, development and working groups, and governance processes across the HEI. This resource is multiplied across all submitting units, which in larger institutions can be substantial: in REF 2021, 31 institutions submitted in 20 or more UOAs. Based on averages alone, the potential reduction in the submission requirement is substantial. The average number of submissions in REF 2021 (12), of average size (40.3 FTE), would have required a maximum of 124,800 words across the unit-level environment templates. This contrasts with just 4,500 for the IL statement for the average-sized institution in the pilot, showing a reduction of 96 per cent. Even with review of the IL word limits, the potential saving would still be significant. The single IL environment approach would also remove an element of the assessment burden from the assessing sub-panels.

Recommendations: Overall

R1. For future research assessment, the environment element should be assessed through a single institutional-level submission. This reflects the clear opportunities for reducing burden for submitting institutions and the high value offered in requiring information about the institution as a whole. It is further supported by the broad alignment observed between the outcomes from the pilot and the average UL environment outcomes. The detailed submission requirements should be refined in consultation with the sector.

R2. The submission should incorporate and enable understanding of contributions and brief key information from each of the submitting units within the HEI.

Submissions

59. Many submissions provided clear, well-evidenced statements that highlighted the vitality and sustainability of the research environment. However, the range of challenges we identified with much of the submitted material suggests the need for some refinements to the guidance and submission requirements.

60. As set out in section 2, a key challenge we identified related to the submission requirements for single-UOA institutions. We noted a particular disparity between those submitting the lengthier REF5b and those small and specialist institutions that chose to submit a REF5a, given the extent of cross-referencing between the REF5a and 5b. We therefore concluded that the most equitable approach for any future exercise was for all participating institutions to follow the same submission requirements. Development of the requirements in consultation with the sector will need to ensure full representation across different HEI types, including small and specialist institutions.

61. The specified word limits for the REF5a statement were scaled according to total number of staff submitted by the institution and were modest for the pilot exercise. In practice, size alone did not always equate to the complexity of some institutions. We therefore propose that the word limits are reviewed for a single IL assessment, including whether additional factors to FTE alone should be considered in relation to the scaling of word limits.

62. Further challenges included inconsistency in evidencing and use of quantitative indicatives; a lack of clear identification of outcomes; failing

to articulate how investments, activities and so on linked to research; and including significant material in the wrong sections of the template – in particular in the non-scored 'Context' section – and in some cases then leaving too little detail elsewhere. While we concluded that the statement was broadly focused on the appropriate features, we recognised that the 'pilot nature' of the exercise had made it difficult for evidence to be delivered effectively.

63. We propose that the identified issues would best be addressed through a more prescriptive template and guidance, provided at an earlier stage, alongside the wider learning that will take place through the pilot outcomes and publication of the statements. We concluded:

    a. There should be an increased focus on institutional governance and recognition for supporting structures and staff, beyond academic and professional services staff.
    b. That a more prescriptive template should aim to ensure the information is outcomes-focused and clear about the relevance to research.
    c. That a more structured template should increase the amount of quantitative data provided, in a standardised format. Examples include consistent requirements for EDI data on appointments and promotions.
    d. To ensure standardisation of the information provided about submitting units as part of the single IL submission, we also concluded that this aspect would need to be closely specified. This should include key data and brief narrative supporting information, outcomes for PGR, retention of ECRs and proportion of fixed term contracts.

64. The statement should not have sections that are not scored to ensure all submitted material can contribute as relevant towards the outcome. Given the overlap often observed in the 'Context and mission' section with the 'Strategy' section, it may be appropriate to incorporate the non-scored 'Context' into 'Strategy'.

65. It will be critical for the detailed requirements to be developed through close engagement with the HE sector. In view of the degree of the changes proposed, this should focus on consultation and development of initial guidance as early as possible.

Recommendations: Submissions

R3. All participating institutions should follow a single format for IL environment submissions.

R4. Work with sector representation should seek to identify the impact of any changes for smaller/less-research-intensive HEIs and consider how these could be addressed.
R5. The submission template and guidance should be more tightly specified, including standardised supporting data. This approach should also be extended to the UL information to be provided as part of the IL submission.
R6. All elements of the submission should be scored; the context section in the current template could move into the strategy section.
R7. Detailed requirements should be developed in consultation with the sector, with an aim to provide initial guidance at an early stage.

Assessment process

66. The assessment process itself worked well overall. Some of the challenges we encountered in applying the criteria would likely be addressed through the development of a more tightly-defined template as covered above.

67. The criteria of vitality and sustainability were appropriate and could be fairly applied in the pilot process. We noted a range of missions across the submissions we reviewed, including institutions that were, or aimed towards, contributing to or leading research in an international context, and those playing very significant roles in national or local contexts. We considered the degree to which the REF concept of excellence – particularly as defined through the descriptions of the starred levels – should be broadened to support a wider recognition of excellence at the highest levels.

68. The holistic framing of the environment level definitions incorporate the highest quality ratings for both outputs and impact in terms of what the environment is judged to be conducive towards producing: i.e., four star means conducive to producing world-leading research and enabling outstanding impact. Through our assessment, we sometimes observed a conflict between excellence in research and enabling its impact, and propose that this issue is considered carefully in defining the criteria of a future exercise

69. The processes we undertook to calibrate our scoring and moderate this throughout the assessment stage were detailed and robust for the purposes of the pilot. We found the sub-group model to be effective in reaching proposed scores for each submission and found the process for agreeing scores as a panel sufficient overall, although somewhat

compressed. To ensure appropriate rigour in the application of standards across groups in a future assessment, we consider there is room to further extend calibration and moderation processes, as well as the time dedicated to final review by the panel collectively.

Recommendations: Assessment

R8. The starred level definitions should be reviewed to support a broad definition of excellence and to consider the relationship between supporting research and enabling impact.

R9. Calibration, moderation, and final review assessment processes should be extended in a future assessment.

Panel

70. The pilot panel comprised a relatively small number of members, although constituted to include a broad spread of academic, senior managerial, research user and research professional expertise, representing a range of institutional types – including small and specialist institutions. In addition, the panel was able to draw on advice from the international members who were appointed at the start of the assessment stage and brought experience from membership of REF 2014 panels. This meant that the panel was able to call upon a range of specialist expertise, as well as ensuring a plurality of views and experiences.

71. We considered the mix of our panellists' expertise was broadly appropriate for the task of assessing the IL environment, while identifying particular areas where this could be extended for a future assessment. This reflects our view on the particular nature of the task involved in assessing the IL environment which differs from the more discipline-focused nature of the task at the level of the submitting unit.

72. Therefore, we agreed that IL assessment should continue to be undertaken by a single panel, representing a range of expertise relevant to assessment of the wider institutional environment – including significant experience of strategic leadership, senior research management and administration, across a mix of UK institution types and international experience. The perspective of those experienced in the use, benefits and outcomes of research will be a vital component, as will that of individuals with key experience of the unique context of small and specialist institutions. A panel comprising this range of experience will

provide appropriate expertise to assess institutional decision-making and the impact of activities at an institutional level.

73. For a future panel, given the potential scale and implications of the assessment, we considered that the membership would need to expand from that of the pilot panel in terms of the number of members and the range of expertise available. The panel noted this should include in particular expertise in research integrity and EDI legislation and practice.

74. As outlined in our overall reflections, we propose the inclusion of specific information at unit-level as part of the IL assessment in future. It would be vital, therefore, for structured and effective processes to be put in place to allow advice to be exchanged between the assessing sub-panels and the IL panel.

Recommendations: Panel

R10. A single IL panel should be appointed to undertake assessment across all submitting institutions.
R11. The IL panel should be constituted to provide expertise at the appropriate level to assess the wider institutional environment, including higher managerial, strategic and administrative, research user, and small and specialist experience.
R12. The panel should be extended to include specific expertise in research integrity issues and EDI legislation and practice.
R13. The assessing panel should work with the sub-panels for advice on disciplinary-relevant elements of the submissions.

Annex A

ILEPP membership

Panel chair		
Professor Chris Day	Newcastle University	

Deputy chair		
Professor Michael Fitzpatrick	Coventry University	

Members		
Professor Alan Barrett	University of Texas Medical Branch	From Jun 2021
Professor John Cattell	Historic England	
Dr Stephen Conway	University of Oxford	
Professor Nandini Das	University of Oxford	
Professor Kirsten Drotner	University of Southern Denmark	From Jun 2021
Professor Sir Barry Ife	Guildhall School of Music and Drama	
Professor Andrew Jones	Brunel University London	
Professor Linda King	Oxford Brookes University	Until May 2022
Professor Fiona Lettice	University of East Anglia	
Professor Dewi Lewis	Independent	
Professor Weiru Liu	University of Bristol	
Professor Ruth Northway	University of South Wales	
Professor Mark Ormerod	Keele University	Until Jan 2022
Professor Murray Pittock	University of Glasgow	
Professor Michael Rayner	University of the Highlands and Islands	
Dr Rosa Scoble	Brunel University London	
Professor Martin Tillotson	University of Leeds	
Mr Alisdair Wotherspoon	Independent	

Observer		
Professor Dianne Berry	University of Reading	

Panel secretary		
Myles Furr	Research England	