

The Corona-Eye: Exploring the risks of COVID-19 on fair assessments of Impact for REF 2021

Authors

Gemma E. Derrick¹

Julie Bayley²

Affiliations

¹ Centre for Higher Education Research & Evaluation, Department of Educational Research, Lancaster University, United Kingdom, LA1 4YD (g.derrick@lancaster.ac.uk)

² Director of Research Impact Development, VC Office, University of Lincoln, United Kingdom (jbayley@lincoln.ac.uk)

Abstract

This paper assesses the risk of two COVID-19 related changes necessary for the expert-review of the REF2021's Impact criterion: the move from F2F to virtual deliberation; and the changing research landscape caused by the COVID-19 crisis requiring an extension of deadlines, and accommodation of COVID-19 related mitigation. Peer review in its basic form requires expert debate, where dissenting opinions and non-verbal cues are absorbed into a groups deliberative practice and therefore inform outcomes. With a move to deliberations in virtual settings, the most likely current outcome for REF2021 evaluations, the extent that negotiation dynamics necessary in F2F evaluations are diminished and how this limits panelists' ability to sensitively assess COVID-19 mitigation statements is questioned. This article explores the nature of, and associated capabilities to undertake, complex decision making in virtual settings around the Impact criterion as well the consequences of COVID-19 on normal Impact trajectories. It examines the risks these changes present for evaluation of the Impact criterion and provides recommendations to offset these risks to enhance discussion and safeguard the legitimacy of evaluation outcomes. This paper is also relevant for evaluation processes of academic criteria that require both a shift to virtual, and/ or guidance of how to sensitively assess the effect of COVID-19 on narratives of individual, group or organisational performance.

Introduction

As the global academic community works to recover and reorganize its research practice during and in anticipation of a 'post' COVID-19 normal, so too do existing evaluation and governance practices need to adjust. Although delayed, the UK's Research Excellence Framework, and the evaluation of the Impact (capitalized to reflect the formal requirement within REF) will proceed, but with alterations to previously set, deadlines; census dates; evaluation practices; and how COVID-19 is taken into account as a reasonable adjustment and alteration to research productivity expectations pre-COVID-19. However, despite extended submission deadlines and census dates for Impact, and accommodation for changed Environments, there is little expressed understanding about how disruptions because of COVID-19 will be evaluated and how the adjustments necessary for panel evaluations to continue will influence this process. This paper considers the effect that changes necessary for a post-COVID-19-normal assessment processes may have on the evaluation of the non-academic, ex-post societal impact (the Impact criterion), as distinct from the two other REF components (Outputs and Environment). More specifically it will explore the connected implications of a) the shift from face-to-face to virtual review processes, and b) the challenges of judging the complex object of impact, now unsteadied further by COVID-19. It is important to explore both the theoretical and conceptual foundations and risk posed to the assessment of Impact by the COVID-19 crisis ahead of future empirical work exploring evaluative practice during the REF2021 evaluation process. The paper concludes with comments relating to potential risks associated with these aspects, and call for transparent expression of how these are to be integrated into the impact assessment in this unprecedented age. It further offers the cautions outlined in this paper to international counterparts embarking on national (or other such significant) impact assessments to enable fair, equitable and transparent processes to be established from the outset.

Impact and formal assessment: a global enterprise

The deepening formality of impact evaluation has accelerated in the last two decades globally and is continuing to scale across nations with the UK's Research Excellence Framework arguably the most dominant expression to date. Germinated but paused by Australia in the early 2000s (and since reinvigorated), the principles of evaluating the non-academic benefits of university research are powerfully rooted on policy aims to demonstrate the contribution of academic research to broader society ((HEFCE) 2009). More specifically impact evaluation seeks to address 4 As - advocacy, analysis, accountability and allocation (Adam et al, 2018), more recently extended to include a further 2 A's (acclaim and adaptation, https://www.rand.org/pubs/research_reports/RR3200.html) - offering a mechanism to not only inform government decision making about funding such as in the case of the UK REF, but also routinise expectations of 'effect' within the research sector. Notwithstanding

1
2
3 broader debates about the drawbacks or merits of peer review compared to metrics-driven
4 approaches (Wilsdon 2016), peer review remains a primary format of assessment,
5 particularly in the UK. However, impact is far from a globally embedded or academically
6 internalized aspect of research life, with considerable international variation in the extent to
7 which an impact agenda exists/is implemented. Furthermore, impact evaluation is
8 complicated by a variety of impact models, frameworks and heuristics available to
9 understand it (Cruz Rivera et al. 2017). There is also criticism that REF is an expensive, over
10 managerialist process that has little benefit for UK research (Watermeyer 2019).
11 Accordingly whilst there is extensive coverage of how impact acts as a demonstration of
12 'public good' from research (Charities 2019) the operationalization of impact evaluation
13 within academia continues to unsteadily walk the tightrope between social responsibility
14 and instrumentalization.
15
16
17
18
19
20
21

22 **Impact assessment is still complex: the need for the Evaluator's Eye**

23 Despite the global implementation of Impact as a formalized evaluation criterion, Impact
24 evaluation is still not a standardized endeavor and multiple models of impact assessment
25 are used even within single domains such as health (Milat, Bauman, and Redman 2015;
26 Greenhalgh et al. 2016; Cruz Rivera et al. 2017). As a result, the criteria and practice of
27 consideration for such assessment is not fixed. Whilst the term 'impact evaluation'
28 conceptually represents the calculable endpoint of research implementation, unlike
29 traditional models of research excellence (such as outputs), there is no clear point at which
30 impact is 'done'. Furthermore, positivist paradigms underpinning causation and
31 measurability within impact assessment are not universally accepted across disciplines
32 which are rooted in more transactional and subjective connections with society (Crossick
33 2016; Stern and Seifert 2016). Thus, research impact evaluation is a complex mix of primed
34 expectations, disciplinary philosophy and multicomponent judgements of non-static objects.
35
36
37
38
39
40
41

42 Beyond models of Impact to understand its generation and aid its evaluation, Derrick (2018)
43 examined how a focus on the practice of impact evaluation generates new models of impact
44 appreciation that are specific to the place, and context of the evaluation. Here, the
45 importance of panels developing an "Eye" or a strong panel identity and working methods
46 was shown as crucial to reach a consensus to evaluate complex objects, such as non-
47 academic, societal impact. Even though this research took place during the 2014 Research
48 Excellence Framework, when Impact assessment was arguably more nascent, reviewers
49 more novice and expectations less socialized, Impact definitions, exclusions and inclusions
50 have remained substantively unchanged. Whilst some aspects have moderately evolved –
51 such as improved accommodation of public engagement and impact on teaching in its
52 REF2021 definition – the overall assessment objects within impact remain ambiguous,
53 complex, resistant to simplistic mechanisms of academic evaluation and therefore in need
54 of increased levels of deliberation within groups such as peer review panels. Within steady
55 times, judgement of impact is difficult enough, but the extraordinary circumstances of
56
57
58
59
60

1
2
3 COVID-19 add challenges and complexities that are previously unexplored. It is therefore
4 imperative that the risk that the new normal posed by the ongoing COVID-19 crisis to the
5 evaluation of Impact be explored theoretically ahead of being tested empirically.
6
7

8 **COVID-19 and the UK Research Excellence Framework 2021**

9
10 COVID-19 has already had a significant effect on the immediate future of
11 universities. Beyond the obvious (and significant) pressure on individuals to operate within
12 a pandemic, there have also been numerous practical and staffing changes. For instance,
13 universities have been routinely using their Polymerase chain reaction (PCR) machines and
14 experienced staff and technicians to support the UK National Health Service (NHS) to deliver
15 COVID testing. This reduces their capacity for non-COVID research, placing further strain on
16 the future of UK university research. Expertise has also been redeployed towards the
17 pandemic response, with such high-profile examples as vaccine development at the
18 University of Oxford. Notwithstanding the implicit value of such resource diversion for
19 public health, the cannibalisation of resources from non-COVID activities compounds the
20 challenges for those already overstretched by, for instance, the time burden to convert
21 teaching into virtual formats. Following a consultation about the REF-specific complications
22 arising from COVID-19, Research England¹ made a series of alterations to the submission
23 process to partially accommodate for these unprecedented pressures. These included a 4-
24 month extension to the final submission date, and guidelines on issues such as inclusion of
25 furloughed staff (point 18), delayed outputs (point 28) and changes in the research
26 environment (point 64). For Impact (but not Outputs), a revised timescale - shifted back
27 from 31st July to 31st December 2020 - effectively provided an additional 5 months for
28 Impacts to develop, and with accommodation for accounts of disruption for resulting Impact
29 Case Studies (ICS). In parallel, re-scheduled panel review processes have shifted to virtual
30 deliberations. Whilst optimism pervades of a potential shift back to physical opportunities
31 for debate, at time of writing, the persistency of COVID-19 means that online methods
32 remain the primary and potentially only method for review.
33
34
35
36
37
38
39
40
41
42
43

44 **Why raise concerns?**

45
46 Such modifications, however practical, introduce altered accounts and new dynamics to the
47 impact evaluation process. Group peer review² in its most basic form requires the presence
48 of experts in a forum where debate is encouraged to achieve consensus and resolve dissent,
49 with new approaches by necessity deviating from more traditional face-to-face (physical co-
50 location) methods. Fundamental to fair decision making in impact assessment is recognition
51
52
53

54
55 ¹ <https://www.ref.ac.uk/media/1417/guidance-on-revisions-to-ref-2021-final.pdf>

56
57 ² For the purposes of this paper the terms: 'groups', 'teams' and 'panels', will be used to refer to the process of
58 peer review conducted by more than one of expert through deliberation.
59
60

1
2
3 that the benefits of evaluative decision-making through peer review are not solely
4 associated with the *nature* of the outcome (Derrick 2018), but also through demonstrations
5 of *deep evaluative inquiry* via a deliberated and quality process. However, not only have
6 review mechanisms needed to shift to virtual platforms, the subject of evaluation - Impact
7 case studies - have themselves been affected by the COVID-19 crisis introducing need to
8 account for changed trajectories and issues in evidencing. Given the financial and
9 reputational weighting of the REF process, and the scale of the peer review machinery to
10 reach judgements about ICS, it is essential that this process be transparent and attentive to
11 the possibilities of bias, however unintentional. There is no criticism of the REF2021 for
12 adapting in the face of COVID-19, but left unchecked these issues can too easily corrode the
13 perceived legitimacy of – and academic trust in - evaluation of Impact. Accordingly it is the
14 *absence of knowledge* of these aspects, rather than a presumption of deficient mechanisms,
15 that underpins this paper.
16
17
18
19
20
21
22

23 To address this, we here draw attention to two separate (but connected) factors central to
24 maintaining the legitimacy of evaluation in this COVID-adjusted context. Firstly, we explore
25 the implications of (A) a move from face-to-face (F2F) to a virtual peer review evaluation,
26 highlighting potential risks of presuming the benefits of physical review automatically
27 transfer to online platforms; and (B) the challenges arising from the extended Impact
28 deadline, both to the nature of impact case studies submitted, and the equitable judgement
29 of impacts differentially affected by the pandemic. For each change we reflect on a series of
30 considerations for the process of Impact peer review, followed by a set of
31 recommendations.
32
33
34
35
36

37 **Change A: The move from F2F to a virtual peer review evaluation**

38
39
40 Panels faced with the evaluation of Impact do so amidst wider political pressure to ensure
41 the ‘right’ outcomes are produced. Achievement of this goal is dependent not only on the
42 content of the ICS, but also the combined understandings, shared experiences, existing
43 collaborative relationships or acquaintances and resultant interactional expertise developed
44 within groups at the time of the assessment (Derrick 2018). The shift to virtual platforms is
45 eminently sensible in the current global context, and the increased efficiency combined with
46 reduced cost makes the adoption of a virtual panel meetings an obvious choice during
47 COVID-19, during which social distancing and restriction of travel prohibits F2F. However,
48 whilst such a shift in normal times would be sufficiently paced to accommodate transitional
49 planning, the ‘new normal’ has hastened the adoption of online infrastructures arguably
50 without opportunity for due diligence on implications for academic governance and
51 evaluation processes. Early indications show panellists reporting a decreased attention
52 span (Zoom fatigue) during virtual peer review panel meetings (NIH Center for Scientific
53 Review 2020; Singh Chawla 2021), and this move away from F2F evaluations further risk
54 evaluations that are dependent on group interactions, such as REF2021.
55
56
57
58
59
60

Consideration 1: Maintaining sufficient peer review deliberation in the virtual setting

Within peer review decision making - where the decision is the collective responsibility of the group - the quality of the outcome is a direct consequence of the quality of the deliberation process between experts, or research peers (Lamont 2009; Derrick 2018). Communication, so central to this process, is in normal times facilitated by physical meetings in which individuals, now collected into a panel, debate and discuss the review object and reach a conclusion. The risk of *decreased* deliberation quality that is of particular concern for the evaluation of Impact which is, due to the increased range of experiences and expertise available to panels during the 2021 REF exercise, is already complex and difficult to achieve a group consensus and outcome. Thus, whilst a shift to virtual communication is undoubtedly the right course, caution is needed to avoid presuming vital deliberative processes are naturally transferred from physical to non-physical arenas.

Previous research on peer review decision making demonstrates an overall superiority of F2F methods. Studies show no significant difference in the quality of the decision making in teams using written (text-only) or audio-only communication (Gallo et al. 2013), but do show a benefit of adding video formats (including videoconferencing facilities) resulting in a significant improvement in the quality of team's deliberations and resultant strategic decision making (Baker 2002). Comparative research on the scoring patterns of peer reviews of traditional academic criteria has found only subtle differences in those between F2F and virtual panels; however, there was a decrease in discussion *quality* when deliberations were conducted virtually (Carpenter et al. 2015; Gallo, Carpenter, and Glisson 2013; Gallo et al. 2020). This is the case regardless of whether virtual deliberation is conducted using video conferencing, instant messages or other supported web technologies (such as parallel chat functions) (Carpenter et al. 2015; Gallo, Carpenter, and Glisson 2013; Gallo et al. 2020; Pier et al. 2017). Thus, regardless of the sophistication of web-technologies to attempt to recreate the benefits associated with F2F, F2F remains the preferred, and more efficient mechanism to support complex decision-making, with this effect amplified for new panel groups (O'Neill et al. 2016).

For peer review panels, F2F communication allows for better integration of expertise across panel members when evaluating as part of a group regardless of the group's history in working together, this integration of outside or minority voices is more difficult to encourage virtually. This is especially the case when non-verbal communication in virtual teams is minimized and yet still plays a large role in recruiting allies and including otherwise silent members during panel deliberations. More so than ever, the non-verbal, sometimes unconscious heuristics are essential for (large) evaluation panels to navigate the complex, and ambiguous object such as Impact, for assessment, and the availability of these cues are likely to be reduced when the deliberation is moved from F2F to a virtual setting.

1
2
3 Ensuring sufficient debate with all voices is a vital mechanism for avoiding groupthink
4 (Derrick 2018), but this is made particularly challenging in virtual settings which denaturalize
5 multiway discourse, limit the full airing of disagreements, and more readily default back to
6 turn-taking monologues (Nemeth 1995; Nemeth and Rogers 1996; Yilmaz and Pena
7 2014). Interestingly, data suggests that differences in the quality of deliberations disappear
8 for experienced reviewers³ (Lam and Schaubroeck 2000; Schaubroeck and Yu 2017a), as
9 long-standing teams are more resilient in virtual settings (Hollenbeck, Beersma, and
10 Schouten 2012; Miles and Hollenbeck 2014), emphasising the importance of panel
11 culture (Maruping and Agarwal 2004; Lamont 2009; Derrick 2018; Kozlowski and Bell
12 2013) in navigating complex evaluation objects. However, given the mix of new and
13 experienced reviewers in REF2021, attention is needed on the effects of the virtual shift on
14 the degree of panel cohesion (“temporal stability”) where panel members are new to
15 working together (Hollenbeck, Beersma, and Schouten 2012), and the extent to which
16 heterogeneity in panel characteristics differentially affects their engagement with virtual
17 methods (Schaubroeck and Yu 2017b).

18
19
20
21
22
23
24
25
26 Part of the practical moves towards more efficient evaluations comes from the explicit or
27 implicit practice of decreasing the amount of deliberation though sidelining dissent or
28 requests for clarification. Indeed, without the human distractions that are inevitable in F2F
29 meetings, is a tendency towards streamlining discussions for the sake of it occurring online
30 and therefore being too quick in discussions and not exploring things fully. In most
31 situations, these additional needs for information are likely to originate from out-group
32 members who are either impact novices with a low temporal stability and members of the
33 interdisciplinary panel (who might create dissent by questioning the consensus otherwise
34 established round disciplinary boundaries). More often than not, membership of these out-
35 groups can perpetuate existing gender, racial or geographical biases in academia thereby
36 limiting the ability for these individuals and groups to inform the evaluation process if there
37 perspective counters the established consensus, and if there is a disproportionate focus on
38 efficiency at the expense of complete, and fair deliberation. With the difficulties associated
39 with considering non-verbal cues via virtual platforms (Lam and Schaubroeck 2000; Miles
40 and Hollenbeck 2014), there is also less information available to the Panel Chair, or
41 individual panelists to gauge when their actions or deliberations are indirectly or unfairly
42 sidelining the opinions and inputs of out-group members, or when the virtual platform is
43 being used to dissuade dissent or avoid conflict that could be used to improve the
44 decision. Therefore, more care is required to ensure that the panel members are engaged
45 with the evaluation process, their perspectives heard, and to detect non-verbal cues from

56
57 ³ Here ‘experience’ refers to previous experience assessing Impact; as well as experience working with other
58 evaluators (resulting in a high degree of temporal stability) as part of the 2014 Research Excellence Framework.
59
60

1
2
3 panelists that indicate dissent from the consensus, but where panelists are dissuaded from
4 raising dissent or needs for clarification verbally.
5
6

7 **Consideration 2: Low versus high temporal stability**

8
9
10 Temporal stability refers to the degree to which team members have a history of working
11 together in the past and expect to be working together in the future (Schaubroeck and Yu
12 2017a; Hollenbeck, Beersma, and Schouten 2012). Teams that have a shared history of
13 working together develop implicit norms and certain familiarities with one another, thereby
14 reducing much of the uncertainty associated with how tasks are completed (Hackman and
15 Katz 2010). Similarly, teams that expect to remain working together in the future are more
16 motivated to invest the time to develop these norms to better facilitate how work is
17 accomplished (Driskell et al 2003; Maruping and Agarwal 2004). Previous research has
18 shown how teams with high degrees of temporal stability make better decisions than those
19 with low temporal stability (Lu, Yuan, and McLeod 2012) and that complex decision making
20 (Schaubroeck and Yu 2017b) including decision making around Impact, is made more
21 difficult in less stable teams.
22
23
24
25
26
27

28 For REF2021, panels will be made up of both REF (and Impact) novices, as well as more
29 experienced evaluators from REF2014, bringing challenges for the 'temporal stability' of
30 panels. This is of course not uncommon within academia, with the frequent convening of
31 new groups to assess applications, promotions and other such standard aspects of academic
32 life. Impact and REF-novices have different deliberative and heuristic needs than evaluators
33 who are either more experienced with the REF evaluation, or with working together.
34 Accommodating these different needs, although minimal in F2F groups, is made more
35 difficult if deliberations are conducted virtually (Driskell et al 2003; Schaubroeck and Yu
36 2017b). This is because panels with a high degree of temporal stability are able to bypass
37 the benefits of F2F communication, since they are already familiar with the characteristics,
38 expertise and potential contribution of each member, can anticipate more salient
39 behavioural messages, the potential meanings of silence and therefore the reliance on the
40 need for clarification (by confirming 'have we decided on that?') is low (Schaubroeck and Yu
41 2017b).
42
43
44
45
46
47
48

49 For the REF-panels, where panelists may already have existing academic relationships (e.g.
50 collaborative, academic or professional) in addition to previous experience of REF-work,
51 there is a need for a balance between; using the high temporal stability of in-groups to
52 minimise the level of deliberation necessary to reach a decision, thereby increasing the
53 'efficiency' of the evaluation; and, promoting a longer, more 'inefficient' deliberation
54 capable for compensating for the low level of temporal stability between in- and out-groups
55 but aiming to allow groups the time necessary to fairly and robustly assess the mitigation
56 within each ICS. Whereas one option provides a quicker, more efficient outcome, it does so
57
58
59
60

1
2
3 at the expense of developing the panel culture necessary to robustly evaluate Impact in a
4 COVID-19 context. The other option may result in a 'better decision' due to the reduction in
5 temporal stability through increased deliberation within the team but does so at the
6 expense of time and efficiency.
7
8
9

10 **Consideration 3: Ensuring team cohesion and culture**

11
12 Teams that have a shared history of working together develop implicit norms and certain
13 familiarities with one another, thereby reducing much of the uncertainty associated with
14 how tasks are completed (Hackman and Katz 2010; Kozlowski and Bell 2013). This panel
15 culture or 'Eye' (Derrick 2018) is an important driver of the evaluation, especially when
16 tasked with complex decision-making. However, a low level of temporal stability (described
17 above) in the first instance, does not imply that it will have an ongoing effect on the
18 development of panel culture throughout the assessment process. Teams with less history
19 of working together and who come together for the completion of a specific task are still
20 able to work effectively together, as well as minimize the disadvantages associated with low
21 temporality teams, but this is only possible if sufficient deliberation time is provided for the
22 necessary development of the psychological linkages within teams including a willingness to
23 contribute to a collective goal (Wiesenfeld, Raghuram, and Garud 2001). Increased and
24 more mindfully managed deliberation, by the Panel Chair for example, would have the
25 advantage of allowing groups to bring in voices and perspectives that would otherwise be
26 sidelined because they originate from peripheral groups (low temporal stability) or that
27 promote dissent in a consensus (judge) decision frame. This process, and the need to reach
28 a consensus representative of all panelists, takes time that goes beyond a discussion that is
29 driven solely by the need to produce an outcome.
30
31
32
33
34
35
36
37
38

39 **Consideration 4: Potential for conflicting decision frames in F2F versus virtual settings**

40
41 Previous research has shown how evaluation groups adopt a "solve" decision frame for
42 aspects of non-academic impact (Derrick 2018) where the decision is framed around
43 choosing the best option, rather than striving to reach a consensus as per a "judge"
44 decision frame (Stasser and Stewart 1992). For the REF2021, the otherwise complex expert-led
45 decision making around Impact is made more complex by the additional evaluation required
46 around the mitigation needs within ICS. This increased complexity, requires more complex
47 evaluation processes and tools to be applied by panel members. Further, whereas F2F
48 teams are able to engage in more complex decision-making via a preferable 'solve' decision
49 frame, to move to virtual deliberations complicates this process, as does the low level of
50 temporal stability within panels. In a 'solve' decision frame, the team tries to find an optimal
51 solution by seeking out information that reaches a demonstrably correct answer, thereby
52 searching (and finding) information that allows the team to identify, and therefore defend,
53 their choice as a decision (Weiss and Bucuvalas 1980; Stasser and Stewart
54 1992). Notwithstanding the implications of confirmation bias in group decision making
55
56
57
58
59
60

1
2
3 (Stasser and Titus 2003), there is clear need to understand the decision on, and processes
4 arising from the adopted decision frame.
5
6

7
8 In addition, achieving a more ideal decision frame during virtual deliberations requires a
9 greater level of complexity. A more time-consuming deliberation would be needed to
10 compensate for the ease by which virtual platforms can silence non-peripheral, or even
11 dissenting voices when the decision frame adopted by the group is seen to act against issues
12 that are in need of more time, consideration and deliberation in order to reach an optimal
13 solution. More likely, teams will tend towards a 'judge' frame which is based on the
14 perception of a consensus and are more easily achieved for virtual settings. However, this
15 goes counter to the solve frame normally achieved in F2F settings, which values processing
16 all available information. Although it is debatable whether this differing decision frame
17 would produce an alternative outcome, it does speak to the reliability of the evaluation
18 process, the perception that the evaluation is fair and complete and therefore the outcomes
19 accepted by the academic community. Therefore, even though solve frames are more
20 harmful in virtual teams, they remain harmful so long as members do not work to
21 compensate for the loss of social, non-verbal cues than support intensified information
22 search and scrutiny (Dennis et al. 1996; Swaab et al. 2012) and therefore a desirable 'solve
23 frame' which is more easily maintained in F2F settings. As such, where the considerations
24 necessary to provide a robust decision are more complex, and where the deliberation is
25 hampered by virtual communication, more work for panels, and panel chairs is required to
26 encourage an effective and robust deliberation.
27
28
29
30
31
32
33
34

35
36 Of concern is research demonstrating that the medium of communication has an influence
37 on this task of decision making (establishing the decision frame). In both virtual and F2F
38 mediums, the decision frame adopted invokes different decision processes (Stasser and
39 Stewart 1992; Stasser, Stewart, and Wittenbaum 1995; Stasser 1999). These differences are
40 larger for more complex decision making, and when the team has uneven experience of
41 working together between members (Stasser 1999). A shift to virtual raises particular query
42 on key aspects of communication related to group's choice on decision frame: the speed in
43 which a message is received (transmission velocity); the speed in which the receiver can
44 obtain clarification of the meaning of a message (immediacy); and the extent to which
45 multiple cues, verbal and non-verbal, are supported by the medium (symbol variety). The
46 reduction in social cues available in virtual settings, compared to F2F, reduces the type of
47 information search and scrutiny during decision-making necessary to support a 'solve' frame
48 (Dennis et al. 1996; Swaab et al. 2012). Previous research has shown that F2F teams in
49 solve frames make better decisions than virtual teams who adopt the same decision frame
50 (O'Neill et al. 2016; Organ and O'Flaherty 2016). This is because with F2F settings, critical
51 debate is welcomed as it allows for the exploration of dissent in building a dominant
52 definition of the criterion through further deliberation (Nemeth 1995; Nemeth and Rogers
53 1996). In contrast, in communication mediums that are low on transmission velocity,
54
55
56
57
58
59
60

1
2
3 immediacy of feedback and symbol variety – such as virtual mediums – reduced non-verbal
4 social cues needed to interpret meaning result in more frequent requests for clarification
5 from team members, as well as critiques of commonly held (group-based) or individual
6 perspectives being seen as hostile. This can more readily lead to displays of defensiveness
7 which interrupt the panel culture necessary for complex decision making (Derrick, 2018),
8 and can interrupt the information seeking necessary to support a ‘solve’ frame as an
9 effective approach to group decision making (Nemeth 1986). Thus, in virtual settings,
10 critical discussion can unintentionally shutdown communications so needed to optimise
11 outcomes and evaluation processes. In addition, the ability to manage conflict is also
12 reduced in virtual teams who adopt solve frames. The question therefore remains as to
13 how to create a virtual environment that balances the need for dissent, allows for the non-
14 visual/verbal cues necessary to mediate misunderstandings stemming from dissent in a way
15 that produces an optimal, and efficient outcome for the evaluation of the Impact criterion.
16
17
18
19
20
21

22 **Change B: Timescale extension for Impact**

23
24 To partially compensate for the disruption caused by COVID-19, the deadline for the Impact
25 criterion was extended from July to December 2020. Notwithstanding the institution-level
26 challenges of managing shifted deadlines, this change was a relatively straightforward
27 adjustment to provide case study authors with more time to realise the Impact previously
28 envisioned pre COVID-19, or at least have ‘breathing room’ to complete within the
29 pressures of the pandemic. Within this however, timescale extensions have introduced two
30 unanticipated dimensions of ICS; accountancy of ‘affected case studies’ (more commonly
31 understood as mitigating circumstances) and the emergence of unanticipated COVID-led
32 ICS.
33

34 This section discusses how the deadline extensions, and the absence of formal guidelines to
35 panel of how to mitigate the COVID-19 effects on impact claims, compounds the existing
36 complexity of decision-making to one where panels are asked to navigate the evaluation of
37 Impact that is already ambiguous in concept, and now ambiguous in time.
38
39
40
41
42
43

44 **Consideration 5: Complexities introduced by mitigation**

45
46 For those existing ICS materially affected in some by COVID-19, institutions were given the
47 opportunity to provide a 100 word ‘affected case study statement’ explaining how either
48 Impact or evidence gathering were compromised. NB for the purposes of this article, from
49 hereon we refer to the action of authoring the influence of COVID-19 as *mitigation*, and the
50 subsequent panel evaluation of this as *mitigation judgement*. Within the context of ICS,
51 mitigation can be best defined as the process by which the influence of COVID is reflected in
52 accounts, more often than not expressing negative consequences to ICS. Guidance on both
53 the extension and mitigation was clear that both were available options, to be used or not
54 at the institution’s discretion.
55
56
57
58
59
60

1
2
3 However, this procedural clarity belies the conceptual complexity of judging and narrating
4 how cases are 'affected'. Revised REF2021 guidance expresses the nature and optional
5 usage of extra time, the format of statements, but stops short of prescribing 'what counts'
6 as legitimate mitigation. Instead, the requirement is for institutions to outline how the
7 impact has been 'significantly affected by COVID-19' where such contextual information is
8 required for panels to fully understand the case. As such, whilst extension is a clear-cut shift
9 of date, the principle of 'affected' cases demands not only that institutions independently
10 determine what counts as *significant*, but then successfully articulate these precise
11 circumstantial detriments of COVID-19 blindly hopeful they have met the otherwise
12 unscripted criteria for panel agreement. The incentives to maximise scores for, and indeed
13 most successfully mitigate for disrupted cases, raises the very plausible chance that – in the
14 absence of absolute guidance to the contrary - authors will have injected linguistic qualifiers
15 into the ICS itself to express deviation from what 'would', 'could' or 'should' have occurred.
16 Some accounts may conceivably now describe Impact as a past modal-verb (modal of lost
17 opportunity, or 'what could have happened if only COVID-19 not occurred'), whilst others
18 continue to account more traditionally and concretely about what has demonstrably
19 changed.

20 Responsibility then shifts to panels who must add appraisals of extenuating factors to the
21 already complex object of Impact. The process by which this will be achieved, or calibrated
22 across cases and panels to ensure parity with unaffected cases, is yet unclear, bringing
23 concerns over equitability and fairness in outcomes.

24 **Consideration 6: The potential risk of differential and divergent ICS Types**

25
26
27 In the absence of empirical evidence regarding the Impact submissions made by HEIs to the
28 REF2021, the following section is presented in terms of 'risk' and 'potential for risk'. By
29 bringing these potential risks to light prior to the evaluation process beginning (September
30 2021), the aim is to aid panels to evaluate submissions fairly, as well as offer HEI a
31 framework with which to interpret outcomes. In addition, the section is offered as a
32 conceptual framework for future studies seeking to analyse how COVID-19 mitigation was
33 operationalized in REF2021 panels.

34
35
36 The extension of both the REF2021 census and submission deadlines, combined with the
37 rapidly shifting UK research landscape in light of COVID-19 in 2020 (Watermeyer et al.
38 2021a); and the widely acknowledged competitiveness of UK HEIs (Watermeyer, 2019), means
39 that it is reasonable to expect that HEIs would engage in changes to their ICS strategies in
40 order to capitalize on their REF outcomes. A recent survey-based study of UK researchers
41 showed an institutional de-valuing of research during the initial phases of the COVID-19
42 pandemic, except in relation to preparing REF2021 submissions (Watermeyer et al. 2021b).
43 As for both individuals and organisations, altering research strategies in order to maximise
44 outcomes from research performance-based audits is well documented (Watermeyer 2019).
45 However, with the uncertainties to ICS development and completion, especially as it

1
2
3 pertains to evidence collection, directly linked to the COVID-19 crisis, there is a risk that HEIs
4 may use freedoms afforded by extended REF2021 census date and submission deadline to
5 re-position final ICS selection and development in order to maximise institutional
6 performance. This potentially creates challenges for panels who, are already striving to
7 operate as normal, in a new, virtual environment.
8
9

10
11 Whereas the extension of the impact window is a pragmatic accommodation aimed to aid
12 submissions (HEIs and academics) in light of a complex and unprecedented circumstance,
13 little guidance has been supplied to panels about how to mitigate the effects of COVID-19
14 within and between ICSs which have been differentially affected by COVID-19. Whilst in
15 principle all cases should be judged on individual merit, this study envisages unanticipated
16 challenges for panels to evaluate 3 divergent types of ICS emerging as a combined result of
17 the COVID-19 pandemic and extended census/submission deadlines. We summarise the
18 basic dimensions of these in Table 1, discussed further below. We recognise the extent to
19 which ICS submissions to the REF2021 represent these divergent types will not be clear until
20 all submitted ICS are published alongside of the REF2021 results⁴, but consider the logical
21 potential for their parallel existence sufficient necessitates attention on the potential
22 evaluative risks faced by REF2021 panels.
23
24
25
26
27

28 Prior to the COVID-19 crisis, all case studies would have been characterised as Type 1, that
29 is, unaffected by a global shutdown. However, COVID-19 potentially initiates two further
30 types. Type 2 cases would be characterised as ‘continuing’, beginning before and running
31 through COVID-19, requiring authors to react to changing (facilitating or inhibiting)
32 circumstances. Type 3 cases in contrast would have *because of* COVID-19, primarily where
33 research has been adopted into strategy, guidance or practice in support of public health.
34 Sourcing corroboration for any Type of ICS is of course more challenging post COVID-19,
35 particularly given the reduced capacity and availability of third-party testifiers, but Type 2 is
36 arguably disproportionately affected by challenges of corroboration due to the changed
37 circumstances *within* the ICS lifetime.
38
39
40
41
42

43 As impacts in Type 1 ICS occur before COVID-19, claimed effects would be untainted by the
44 change in global circumstances, and thus evidence is (notwithstanding the difficulties
45 obtaining evidence generally) a straightforward corroboration of ‘what happened’. The
46 urgency and severity of the pandemic has led policy makers to draw on research with
47 unprecedented expedience; uptake timelags so routinely expected in the translational cycle
48 (Hanney et al. 2015; Morris, Wooding, and Grant 2011) have been vastly truncated for
49 COVID-relevant research, resulting in the rapid acceleration of research usage with stronger
50 effects than may have been originally envisaged. Accordingly, Type 3 ICS would have the
51 advantage of accelerated effects with the prospect of far more real-time evidence. The
52
53
54
55
56

57
58 ⁴ A database of ICS submissions was established in light of the REF2014 exercise
59 (<https://impact.ref.ac.uk/casestudies/>) and it is expected that the same will occur for the REF2021 submissions.
60

1
2
3 emergence of Type 3 is in line with acknowledged models of evidence-informed policy that
4 reply on the sudden appearance of a social and economic need (Nutley and Webb 2000),
5 pragmatic decisions shaped by political circumstance (Lomas 1997), or else streams
6 converging (Kingdon 1984). Such effects could be found within any Unit of Assessment but
7 could be most reasonably expected within health and medical related disciplines (eg. Main
8 Panel A, UoAs 1-5), with key examples being the preponderance of public health modelling
9 (e.g (Adam et al. 2018)) vaccine development, and initiatives around public behaviour and
10 compliance.

11
12
13
14
15 However, Type 2 ICS would inherit a unique difficulty for REF2021 evaluation panels in
16 comparison to the relative straightforwardness of Types 1 and 3. Revised REF guidance
17 recognises that previously expected access or evidence may become unavailable due to
18 COVID-19, and offers a mitigation mechanism for such cases. However, for those cases
19 running through the pandemic, COVID-19 injects *corroboration uncertainty*, ie. The difficulty
20 in judging if claimed effects post-COVID are substantiated by evidence collected pre-COVID,
21 or proxy (or even absent) evidence, and/or the difficulty for evaluators to estimate the
22 counterfactual against impact claims within ICS. Impact demonstrated pre-COVID-19 may
23 continue along a planned trajectory (for example, continuing to increase), plateau (stall) or,
24 depending on the nature of impact, conceivably be lost (ie. The planned opportunity, event,
25 person or other activity on which the original impact trajectory was dependent was no
26 longer available/possible, preventing the impact materialising in time for REF2021
27 submissions). For example, if an intervention was launched in the health service pre-
28 COVID, with evidence showing patient benefit and service commitment to continued use,
29 but healthcare staff become subsequently unavailable to comment due to the pandemic, to
30 what extent could continued impact be presumed in the absence of further data? To what
31 extent is evidence substantiating pre-COVID effects sufficiently corroborative of continuing
32 or altered impact post-COVID? And in all cases where evidence is no longer available, how
33 would less perfect 'proxy' measures or the overall absence of data be assessed? The
34 existence of these possibilities demands new or alternative accounting, with associated
35 mitigation judgment tools and adapted panellist behaviours to ensure fair and equitable
36 assessment outcomes.

37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
A notable consequence of the emergence of Type 3 is the opportunity for an institution to
choose to submit a new ICS in place of one which has matured over a number of years (Type
1 or 2). This brings implications for staff engagement, recognition and buy-in to the broader
research strategy. Ultimately the REF2021 process offers a route to financial security, and
thus demands institutions optimize their submissions, this is therefore a competitive
decision where COVID-specific case studies are perceived as having a greater potential to
achieve a 4-star rating, by appealing to the "Corona-eyes" of panel evaluators (discussed
below).

Acknowledging individual experiences of COVID-19 - 'The same sea but different boats'

1
2
3 A key related point for judging variations to ICS is that whilst COVID-19 has been a global
4 phenomenon, effects on individuals vary wildly, with differential pressures by gender
5 (Myers et al. 2020), employment sector, personal circumstances (such as caregiving),
6 organisational decisions such as furloughing and many more. It is uncertain therefore how
7 these heterogenous individual panel members experiences of COVID-19 may manifest in the
8 review process, or if these will introduce personal assumptions into judgements of
9 reasonable (versus disingenuous) accounts of disruption. A natural counterbalance to such
10 biases is the development of a robust combined lens (“Corona-eye”) of the group as a
11 whole, able to more collectively assess impact claims and exercise fair evaluative caution in
12 judging the epistemic legitimacy of claimed mitigation.
13
14
15
16
17
18

19 **The combined implications of a shift to virtual and changed Impact scope**

20 The judgement of ICS, and any associated mitigation in reference to COVID-19 is therefore
21 dependent on a mixture of how cases (and mitigations) are articulated, the types of ICS
22 submitted, how mitigation judgement is fostered and calibrated within panels, and how well
23 institutional presumptions of ‘what counts’ match panel schemas. The joint potential for
24 variation in authors approaches to narrating mitigation (Type 2), panel members’ subjective
25 judgments (*mitigation differential*) and unscripted accommodation of the three Types
26 injects concern over parity of outcomes. The unprecedented global scale of COVID-19, and
27 permeation into all aspects of life, risks elevating the perceived significance of work
28 directly relating to the pandemic (Type 3) at the expense of comparatively
29 downplaying unrelated research (Type 1 and/or 2).
30
31
32
33
34
35

36 Even with the presence of Impact evaluators from the previous REF2014, differences in
37 evaluator experiences and definitions of ‘excellence in impact’, alongside different
38 evaluation mechanics for REF 2021 (Derrick 2018) means there is a risk that group
39 deliberation may continue to rely on the establishment of cohesive team behaviours to
40 navigate the complexities of impact. Processes must now take account of conditional
41 impact pathways and differentially affected Types of case, requiring the execution of
42 different practices of mitigation. Whereas *normally*, F2F methods should offer the
43 necessary conditions to address such complexity, there is little clarity on how virtual panels
44 will be fully able to exercise the necessary agility in evaluative practice within the structures
45 of virtual meetings.
46
47
48
49
50
51

52 **Aligning the panels: “Corona-eyes”**

53 The emergence of different impact Types, evaluated in altered fora, and with the high
54 likelihood for incomplete substantiation or terminological inexactitude (‘should have’,
55 ‘would have’) underscore the need to ensure panel processes are sensitive to, and
56 transparent about the accommodation of, new complexities. Maximum mitigation is
57 perhaps to be expected from institutions wanting to capitalize on an otherwise difficult
58
59
60

2020 by securing an equal (or greater) share of REF2021 rewards, this makes decision-making further complex through the need to judge the authenticity of such accounts. Indeed, appropriate, fair and harmonized evaluation of these issues requires panels to not only judge mitigation, but assess within ICS effectively, assessing the legitimacy of institutional claims about trajectory/disruption, whilst simultaneously separating screening out personal experience of the pandemic. We label this ability to judge impact according to pre-COVID best practice, whilst fairly and objectively accommodating the complexities arising from the pandemic and judging authenticity and proportionality of mitigation accounts as “*Corona-eyes*”. We argue that panels cannot fully develop *Corona-eyes* without fully surfacing, and taking steps to address the challenges outlined here. It should be noted that whilst *Corona-eyes* is derived from considerations for REF2021, we believe it is applicable too to funding agencies and promotion panels in accommodating C.V. gaps over the forthcoming years, where the full effect of COVID-19 on individual careers and knowledge production are likely to manifest more profoundly.

RECOMMENDATIONS

To partially remedy these potential imbalances, we summarise here a number of recommendations to support efficient and effective virtual evaluations, particularly in the assessment of complex objects (Impact). These recommendations are also summarised in Table 2.

1. Selecting an appropriate platform for panel deliberation

It is widely accepted that there is no, one technology that can support all stages of a decision-making process, nor replace all the benefits associated with F2F meetings. However, during the COVID-crisis, technology has adapted to the everyday needs of a larger-than-normal population, narrowing the differences between F2F and virtual decision-making. Whereas it is not the appropriate for this article to promote a particular platform over another, there are some characteristics of such platforms that will facilitate the evaluation of Impact using virtual mediums, while allowing panels the flexibility necessary to sensitively mitigate between different type of ICS resulting from the COVID-19 crisis. Where the F2F option is no longer available, agile and flexible teams that - when in coordination with a communication medium that is also agile and flexible - ensure efficient and high-quality decision-making in complex situations (O'Neill et al. 2016).

Any chosen platform therefore must enable sufficient flexibility of members to allow swift and clear communication, and emphasise resolution support. Platforms with asynchronous technologies or group decision support software (GDSS) best mimic F2F interactions, and commonly used approaches such as separation into smaller groups would also help guard against the isolation of any group members and management of large panels. Chat functions may also serve as a useful way to include more voices into these less naturalistic

ways of communicating, and ensure technology does not reinforce unchecked dominant consensus. In addition, in line with the need to instil a desirable level of trust and engagement with the evaluation outcomes, the platform will also need to allow for a high level of security and confidentiality for evaluation deliberations, as well as the characteristics of the submissions.

2. Additional calibration to accommodate mitigation and virtual decision-making

The use of calibration exercises prior to the formal evaluation taking place within panels (Derrick and Samuel 2017), was used successfully during REF2014 as a mechanism to assist robust discussion around Impact, as well as provide an opportunity for panel members to clarify expectations and form a common lens to guide the impact evaluation (Derrick 2018). Calibration exercises, especially when the evaluation is anticipated as more complex, as was the case of Impact in REF2014, are used as an exercise in maintaining consistency and fairness in evaluation. REF2021 panels will of course undertake similar calibration exercises, particularly valuable for ensuring temporal stability and panel cohesion, but it is essential that additional calibration also attends to the combined challenges of virtual and complexified Impact. Here, it is vital that panellists be provided with sufficient detail and training to ensure deliberative practices are sensitive to the vulnerabilities of each type of ICS emerging as a result of COVID-19, as well as aware of how unconscious individual cognitive and emotional biases that may otherwise shadow how COVID-19 mitigation considered towards evaluation outcomes.

3. Structured breakouts and/or discussion times

It is unsurprising that one of the simplest strategies for enhancing deliberation is to enable discussion in within smaller groups. REF2021 panels are exceptionally large, while essential to ensure that the UoA has access the expertise necessary to evaluate all submitted Outputs, ICS and Environment statements, as well as the reap benefits associated with including international evaluators and non-academic experts, may prove burdensome when deliberations take place virtually. There are advantages, therefore associated with splitting the evaluation deliberations around Impact into smaller groups resulting in a micro-panel culture that is able to interact more through an online platform. Smaller groups are also able to more effectively shape their deliberation and decision-making heuristics around non-verbal clues that signal dissent, as well as be able to resolve dissent more openly than otherwise would be possible in a larger panel operating virtually. This indicates that smaller virtual sub-groups are able to exercise a larger dashboard of heuristics available to consider mitigation than for large virtual panels. Although it is not recommended that these smaller, sub-groups work autonomously from the larger panel, it presents a reasonable practice for the initial stages of the evaluation with, working groups then feeding outcomes and processes back to the main group at a later stage. Cross referencing of scores between

1
2
3 these smaller groups, as well as utilising an ongoing form of calibration could also be used to
4 ensure continuity of practice across smaller panels, UoA and Main Panels.
5
6

7
8 If utilising smaller sub-groups is not possible and a real-time deliberation desirable, using
9 virtual platform in a similar way to how F2F operates risks isolating out-group members, and
10 otherwise skewing the evaluation of Impact, and the consideration of COVID-19 mitigation
11 towards otherwise dominant in-panel discourses. With this in mind, another
12 recommendation is to host structured discussion times, (i) between panellists; (ii) between
13 ICS under consideration; and (iii) around emerging mitigation characteristics underpinning
14 the type of ICS emerging as a result of COVID-19. Whereas this technique would ensure
15 that all voices are heard and to move towards the dissolution of out-groups to increase the
16 level of temporal stability within the panel and move towards an inclusive consensus, it
17 would also take a significantly greater amount of time that would contradict the desire for
18 efficiency. In addition, managing a process that would involve all-inclusive deliberation
19 through structured discussions, would also require a larger role for the Panel Chair
20 (discussed below).
21
22
23
24
25
26

27 **4. Maximising the role of the Chair**

28
29 Managing group-peer review, from the perspective of the panel Chair who acts both as part
30 of the panel's collective identity, but also mitigates the process from a top-down
31 perspective, is an aspect of panel peer review operations that is often overlooked. In a post
32 COVID-19 REF2021 the challenge for the Chair will to ensure the smooth virtual deliberation
33 process but also to ensure that during these evaluations the move from F2F to virtual
34 deliberations does not impede the ability of the panel to apply a responsible degree of
35 mitigating judgement to the COVID-19 mitigation statements. This change places a greater
36 responsibility on the Panel Chair as well as a greater level of Chair-awareness of the issues
37 associated with evaluating Impact post-COVID-19.
38
39
40
41
42

43 Where the spontaneity of communication is enhanced in F2F, its absence, combined with
44 the absence on non-vocal/visual cues within groups, restricts the efficiency of virtual
45 teams. Ultimately it is the responsibility of the Chair to ensure the legitimacy of the panel's
46 working methods, as well as the validity of its outcomes, and this presents extra challenges
47 for panel Chairs when this process is undertaken virtually. The role of the chair also has a
48 strong influence on interpersonal team dynamics and trust, which when combined with
49 factors such as explicit management and an awareness of colleagues and their contexts; are
50 essential factors in the efficient operation of virtual teams (Olson and Olson
51 2006). Managing deliberation over virtual settings can also be increasingly difficult for large,
52 diverse teams, especially when effective leadership is highly dependent on quality
53 interactions that are more difficult to establish, and to maintain virtually (Olson and Olson
54 2006). In virtual settings, it is also difficult to avoid hierarchical management styles that
55 negate the chance for all voices to be heard, and regulating inter-dependencies between
56
57
58
59
60

1
2
3 resources, task components and personnel. Indeed, centralized authority has a negative
4 influence of team innovation, learning, adaptation and performance (Schaubroeck and Yu
5 2017a). The alternative, de-centralized authority, especially in the case of large teams (such
6 as the REF2021 UoA Panels), is difficult to maintain and can further impede the temporal
7 stability and construction of the mental modes necessary if this is conducted virtually. The
8 challenge for the Chair, therefore, is to adopt a management style that promotes the
9 inter/intra-personal team dynamics and levels of trust necessary to establish common
10 ground, common conceptual frameworks (“Eye” (Derrick 2018)) within the time allowed,
11 therefore ensuring an efficient panel deliberation process.
12
13
14
15
16

17 Whereas virtual evaluation panels may present the opportunity for a more efficient
18 evaluation overall, minimising the effect of drawbacks associated with the move away from
19 F2F deliberations requires a more dedicated management strategy within the panel.
20 Administering this strategy, requires a larger than anticipated role for Panel Chairs in order
21 to establish and maintain cohesion throughout the assessment process, and to enhance
22 team interactions while ensuring that communication is fluid and not restrictive to any team
23 members during the deliberative and scoring stages. Not only will the Panel Chair need to
24 work harder, monitoring what is said, but who says it but they will also need to do so in an
25 online environment that may prove challenging to monitor parallel lines of communication
26 (the deliberative practice, alongside the chat-function), remain sensitive to non-verbal cues
27 where visibility of the entire panel is limited, as well as to attribute deliberative points to
28 individual panel members, and to silence overly dominant discourses. In light if this
29 increased role, Panel Charis must be given the autonomy to adopt evaluation practices that
30 go beyond the REF Panel guidelines put in place prior to the evaluation in place. In other
31 words, a maximised role of Panel Chair must also come with the power to adapt evaluation
32 processes in practice as difficulties emerge.
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

DISCUSSION

This article has discussed the risks of a post- (or mid-) COVID-19 world to the peer review evaluation of the non-academic, societal impact (Impact) criterion under the UK's 2021 Research Excellence Framework. These risks reflect both factors that have underpinned the nature of the changing research landscape left over from an all but sudden halt to research and research impact activities as a result of the UK's COVID-19 lockdown in 2020; the altered rules and procedures for its evaluation as part of the UK REF2021; and the need to adopt alternative evaluation processes in a COVID-19 safe manner (i.e. the move from F2F to Virtual evaluations). We have sought to foreground the need to address the individual and combined effects of these issues, and have raised concerns about group decision making on a now even more complex object. More specifically, we wish to raise a dual call to ensure panels develop "Corona-eyes" to offset a range of potential risks to fair evaluation introduced by these new conditions, and to do so both ahead of data and alongside live monitoring of REF2021 evaluation processes to empirically examine these issues in practice.

The heterogeneity of Impact requires sector wide and multi-level impact literacy (Bayley and Phipps 2019a, 2019b), not only of the nature of research led change, but also the conceptual, operational and individual aspects which influence the curation and comparative appraisal of cases. Impact is a complex aspect of the research landscape, with formal judgement of it a chronologically more nascent skill than for parallel processes to judge academic outputs. Whilst impact shares many of the challenges of research assessment as a whole – for example the burden on institutions and the reductionism of metrics as proxies for 'Excellence' – it has unique characteristics which make it particularly vulnerable to the shift necessitated by COVID-19. The purpose of this paper was to underscore the interconnected risks facing panels undertaking impact evaluation in these circumstances, and provide insights necessary for appropriate pragmatic and governance modifications to maintain a panel's ability to fully exercise necessary reflexivity.

Most fundamentally, this article has highlighted the potential difficulties in adopting a blind *process-as-usual* approach to the peer review of the UK REF2021's Impact criterion. In addition, any hybrid approach (F2F and online) must further be assessed for any imbalances introduced by a mixed evaluative format, as well as the effects realized by the COVID-19 crisis. This article does not presume insouciance from REF or those undertaking similar evaluations, nor suggest that a shift to virtual is inherently negative. Indeed, virtual decision making processes are more inclusive of not only international voices, but those for whom physical attendance is more difficult or even prohibitive. Arguably therefore there is an advantage to the legitimacy and fairness of outcomes associated with a peer review process. Instead it presupposes that the pace of change in such unprecedented times may easily lead the implications outlined here to be overlooked, and the risk to the legitimacy of the evaluation outcomes, amplified.

1
2
3 In the absence of a global post pandemic reset, peer review is more likely to abandon the
4 traditionality of the F2F structure and embrace the increased familiarity of a system
5 facilitated partially or wholly online. The points raised in this paper are therefore applicable
6 beyond the REF context, and offered as broader reflections for those transitioning to online
7 evaluation processes, or needing to consider how to mitigate judgements of academic
8 performance in light of the COVID-19 crisis. Whilst these concerns are primarily relevant to
9 the UK REF process at this time, the seismic repercussions of COVID-19 mean that the
10 implications for legitimate outcomes not only resonate across the research ecosystem now,
11 but in future performance assessments.
12
13
14
15
16

17 **CONCLUSION**

18 This is an unprecedented time, with unprecedented challenges across all areas of life. For
19 impact assessment the challenges laid out here, and the associated suggestions for redress,
20 offer an opportunity to establish a transparent, fair and robust process for evaluating
21 impact and impact that *'should have been'*. Ultimately no approach is risk free, but it is in
22 the remit of governing bodies to determine how these can be managed in practice, or more
23 specifically develop "Corona-eyes" to offset them. Such practices are vital to retain
24 community trust not only in the evaluative process, but in legitimisation of the resulting
25 financially and reputationally weighted outcomes
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

REFERENCES

- (HEFCE), Higher Education Funding Council for England. 2009. "Research excellence framework: second consultation on the assessment and funding of research." In.
- Adam, P., P.V. Ovseiko, J. Grant, K.E. Graham, O.F. Boukhris, A.M. Dowd, G.V. Balling, R.N. Christensen, A. Pollitt, M. Taylor, and O. Sued. 2018. 'ISRIA statement: ten-point guidelines for an effective process of research impact assessment. ', *Health research policy and systems*, 16: 1-16.
- Ahlburg, D.A. 2020. 'COVID-19 and UK Universities', *The Political Quarterly*, 91: 649-54.
- Baker, G. 2002. 'The Effects of Synchronous Collaborative Technologies on Decision Making: A Study of Virtual Teams', *Information Resources Management Journal (IRMJ)*, 15: 15.
- Bayley, J.E., and D. Phipps. 2019a. 'Building the concept of research impact literacy', *Evidence & Policy: A Journal of Research, Debate and Practice*, 15: 597-606.
- Bayley J and Phipps D. 2019b Extending the concept of research impact literacy: levels of literacy, institutional role and ethical considerations [version 2; peer review: 2 approved]. *Emerald Open Research* , 1:14 (<https://doi.org/10.35241/emeraldopenres.13140.2>)
- Carpenter, A.S., J.H. Sullivan, A. Deshmukh, S.R. Glisson, and S.A. Gallo. 2015. 'A retrospective analysis of the effect of discussion in teleconference and face-to-face scientific peer-review panels', *BMJ Open*, 5: e009138.
- Charities, Association of Medical Research. 2019. "Making a difference: Impact Report 2019." In.
- Crossick, G. 2016. 'Monographs and open access.', *Insights*, 29.
- Cruz Rivera, S., D.G. Kyte, O.L. Aiyegbusi, T.J. Keeley, and M. Calvert. 2017. ' Assessing the impact of healthcare research: a systematic review of methodological frameworks.', *PLoS Medicine*, 14: e1002370.
- Dennis, A.R., J.S. Valacich, T. Connolly, and B.E. Wyne. 1996. 'Process structuring in electronic brainstorming', *Information Systems Research*, 7: 268-77.
- Derrick, G.E. 2018. *The Evaluators' Eye: Impact assessment and academic peer review* (Palgrave Macmillan: London, Uk.).

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Derrick, G.E., and G. Samuel. 2017. 'The future of societal impact assessment using peer review: pre-evaluation training, consensus building and inter-reviewer reliability', *Palgrave Communications*, 3: 17040.
- Driskell, J.E., Radtke, P.H., & Salas, E. (2003). Virtual Teams: Effects of Technological Mediation on team Performance. *Group Dynamics: Theory, Research, and Practice*, 7(4), 297-323. <https://doi.org/10.1037/1089-2699.7.4.297>
- Donovan, C. 2017. 'For ethical 'impactology'', *Journal of Responsible Innovation*: 78-83.
- Gallo, S.A., A.S. Carpenter, and S.R. Glisson. 2013. 'Teleconference versus Face-to-Face Scientific Peer Review of Grant Application: Effects on Review Outcomes', *PLOS ONE*, 8: e71693.
- Gallo, S.A., K.B. Schmalig, L.A. Thompson, and S.R. Glisson. 2020. 'Grant reviewer perceptions of the quality, effectiveness, and influence of panel discussion', *Research Integrity and Peer Review*, 5.
- Greenhalgh, T., C. Jackson, S. Shaw, and T. Janamian. 2016. 'Achieving research impact through co-creation in community-based health services: literature review and case study. ', *The Milbank Quarterly*, 94: 392-429.
- Hackman, J.R., and N. Katz. 2010. 'Group behaviour and performance.' in S.T. Fiske, D.T. Gilbert and G. Lindzey (eds.), *Handbook of social psychology* (John Wiley & Sons, Inc: Hoboken, NJ).
- Hanney, S.R., S. Castle-Clarke, J. Grant, S. Guthrie, C. Henshall, J. Mestre-Ferrandiz, M. Pistollato, A. Pollitt, J. Sussex, and S. Wooding. 2015. 'How long does biomedical research take? Studying the time taken between biomedical and health research and its translation into products, policy, and practice', *Health Research Policy Systems*, 13: 1.
- Hollenbeck, J.R., B. Beersma, and M.E. Schouten. 2012. 'Beyond team types and taxonomies: A dimensional scaling conceptualization for team description', *Academy of Management Review*, 37: 82-106.
- Kingdon, J.W. 1984. *Agendas, alternatives and public policies*, Boston, MA: Little Brown.
- Kozlowski, S.W.J., and B.S. Bell. 2013. 'Workgroup and teams in organizations.' in N.W. Schmitt and S. Highhouse (eds.), *Handbook of industrial and organizational psychology* (John Wiley & Sons, Inc.: Hoboken, NJ.).

- 1
2
3 Lam, S.S.K., and J. Schaubroeck. 2000. 'Improving group decisions by better pooling
4 information: A comparative advantage of group decision support systems.', *Journal*
5 *of Applied Psychology*, 85: 565-73.
6
7
8
9 Lamont, M. 2009. *How Professors think: Inside the curious world of academic judgment*
10 (Harvard University Press: Cambridge, MA.).
11
12 Lomas, J. 1997. *Improving research dissemination and uptake in the health sector: Beyond*
13 *the sound of one hand clapping*, Policy Commentary C97-1, Hamilton, Ontario:
14 McMaster University, Centre for Health Economics and Policy Analysis.
15
16
17
18 Lu, L., Y.C. Yuan, and P.L. McLeod. 2012. 'Twenty-five years of hidden profiles in group
19 decision making: A meta-analysis', *Personality and Social Psychology Review*, 16: 54-
20 75.
21
22
23
24 Maruping, L.M., and R. Agarwal. 2004. 'Managing team interpersonal processes through
25 technology: A task-technology fit perspective', *Journal of Applied Psychology*, 89:
26 975-90.
27
28
29 Milat, A.J., A.E. Bauman, and S. Redman. 2015. 'A narrative review of research impact
30 assessment models and methods', *Health Research Policy and Systems*, 13: 18.
31
32
33 Miles, J., and J.R. Hollenbeck. 2014. 'Teams and technology.' in M. Coovert and L.F.
34 Thompson (eds.), *Psychology of workplace technology* (Routledge: New York).
35
36
37
38 Morris, Z.S., S. Wooding, and J. Grant. 2011. 'The answer is 17 years, what is the question:
39 understanding time lags in translational research.', *Journal of the Royal Society of*
40 *Medicine*, 104: 510-20.
41
42
43
44 Myers, K.R., W.Y. Tham, Y. Yin, N. Cohodes, J.G. Thursby, M.C. Thursby, P. Schiffer, J.T.
45 Walsh, K.R. Lakhani, and D. Wang. 2020. 'Unequal effects of the COVID-19 pandemic
46 on scientists.', *Nature human behaviour*, 4: 880-83.
47
48
49
50 National Institute for Health (NIH) Center for Scientific Review (2020). Impact of Zoom
51 format on CSR review meetings. Internal report. Accessed 14 June 2021:
52 [https://stage.public.csr.nih.gov/sites/default/files/2020-
53 10/CSR%20Analysis%20of%20Zoom%20in%20Review%20Oct%202020.pdf](https://stage.public.csr.nih.gov/sites/default/files/2020-10/CSR%20Analysis%20of%20Zoom%20in%20Review%20Oct%202020.pdf)
54
55
56
57 Nemeth, C., and J. Rogers. 1996. 'Dissent and the search for information', *British Journal of*
58 *Social Psychology*, 35.
59
60 Nemeth, C.J. 1986. 'Differential contributions of majority and minority influence',
Psychological Review, 93: 23-32.

- 1
2
3 ———. 1995. 'Dissent as driving cognition, attitudes and judgments.', *Social Cognition*, 13:
4 273-91.
5
6 Nordback, E.S. 2013. "The influence of emergent technologies on decision-making processes
7 in virtual teams." In *AMCIS 2013*. Chicago, USA.
8
9 Nutley, S.M. and Webb, J. 2000. '*Evidence and the policy process*', in H.T.O. Davies,
10 S.M.Nutley and P.C. Smith (eds) *What works? Evidence-based policy and practice in*
11 *public services*, Bristol: The Policy Press, pp13-42.
12
13 O'Neill, T.A., S.E. Hancock, K. Zivkov, N.L. Larson, and S.J. Law. 2016. 'Team Decision Making
14 in Virtual and Face-to-Face Environments', *Group Decision and Negotiation*, 25: 995-
15 1020.
16
17 Olson, J.S., and G.M. Olson. 2006. 'Bridging distance: empirical studies of distributed teams',
18 *Human-computer interaction in management information systems*, 2: 27-30.
19
20 Organ, D., and B. O'Flaherty. 2016. 'Intuitive decision-making and deep level diversity in
21 entrepreneurial ICT teams', *Journal of Decision Systems*, 25: 421-35.
22
23 Pier, E.L., J. Raclaw, A. Kaatz, M. Brauer, M. Carnes, M.J. Nathan, and C.E. Ford. 2017. "Your
24 comments are meaner than your score': score calibration talk influences intra- and
25 inter-panel variability during scientific grant peer review', *Research Evaluation*, 26: 1-
26 14.
27
28 Schaubroeck, J.M., and A. Yu. 2017a. 'When does virtuality help or hinder teams? Core team
29 characteristics as contingency factors', *Human Resource Management Review*, 27:
30 635-47.
31
32 ———. 2017b. 'When does virtuality help or hinder teams? Core team characteristics as
33 contingency factors.', *Human resource management review*, 27: 635-47.
34
35 Sivunen, A., and E. Nordback. 2014. 'Social presence as a multi-dimensional group construct
36 in 3D virtual environments', *Journal of Computer-Mediated Communication*, 20: 19-
37 36.
38
39 Stasser, G. 1999. 'The uncertain role of unshared information in collective choice.' in L.L.
40 Thompson, J.M. Levine and D.M. Messick (eds.), *Shared cognition in organizations:*
41 *The management of knowledge* (Lawrence Erlbaum Associates, Inc.: Hillsdale, NJ).
42
43 Stasser, G., and D. Stewart. 1992. 'Discovery of hidden profiles by decision-making groups:
44 Solving a problem versus making a judgment', *Journal of Personality and Social*
45 *Psychology*, 63: 426-34.
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 Stasser, G., D.D. Stewart, and G.M. Wittenbaum. 1995. 'Expert roles and information
4 exchange during discussion: The importance of knowing who knows what.', *Journal*
5 *of Experimental Social Psychology*, 31: 31-244.
6
7
8 Stasser, G., and W. Titus. 2003. 'Hidden profiles: a brief history', *Psychological Inquiry*, 14:
9 304-13.
10
11
12 Stern, M.J., and S.C. Seifert. 2016. 'Understanding the value of arts & culture: the AHRC
13 cultural value project (2016) by Geoffrey Crossick and Patrycja Kaszynska,' *Cultural*
14 *Trends*, 25: 277-79.
15
16
17 Stern, N. 2016. "Building on success and learning from experience: An independent review
18 of the Research Excellence Framework." In, edited by Energy & Industrial Strategy
19 Department for Business.
20
21
22
23 Swaab, R.I., A.D. Galinsky, V. Medvec, and D.A. Diermeier. 2012. 'The communication
24 orientation model: explaining the diverse effects of sight, sound and synchronicity
25 on negotiation and group decision-making outcomes.', *Personal Soc Psychol Review*,
26 16: 25-53.
27
28
29
30 Watermeyer, R. 2019. *Competitive accountability in academic life: the struggle for social*
31 *impact and public legitimacy* (Edward Elgar Publishing).
32
33
34 Watermeyer, R., Crick, T., Knight, C., Goodall, J. 2021a. 'COVID-19 and digital disruption in
35 UK universities: Affiliations and affordances of emergency online migration.', *Higher*
36 *Education*, 81: 623-641.
37
38
39
40 Watermeyer, R., Shankar, K., Crick, T., Knight, C., McGaughey, F., Hardman, J., Suri, V.R.,
41 Chung, R., Phelan, D. 2021b. 'Pandemia': A reckoning of UK universities' corporate
42 response to COVID-19 and its academic fallout', *British Journal of Sociology of*
43 *Education*, pp1-16.
44
45
46
47 Weiss, C.H., and M.J. Bucuvalas. 1980. 'Truth Tests and Utility Tests: Decision-Makers'
48 Frames of Reference for Social Science Research', *American Sociological Review*, 45:
49 302-13.
50
51
52
53 Wiesenfeld, B.M., S. Raghuram, and R. Garud. 2001. 'Organizational identification among
54 virtual workers: The role of need for affiliation and perceived work-based social
55 support', *Journal of Management*, 27: 213-29.
56
57
58
59
60 Wilsdon, J. 2016. *The metric tide: Independent review of the role of metrics in research*
assessment and management . (Sage.).

1
2
3 Yilmaz, G., and J. Pena. 2014. 'The influence of social categories and interpersonal behaviors
4 on future intentions and attitudes to form subgroups in virtual teams.',
5
6 *Communication Research*, 41: 333-52.
7

8 Zaharie, M.A., and C.L. Osoian. 2016. 'Peer review motivation frames: A qualitative
9 approach', *European Management Journal*, 34: 69-79.
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1: Three types of Impact Case Study arising from COVID-19

	Description
Type 1	ICS materially <u>unaffected</u> by COVID-19 (e.g. completed before the pandemic)
Type 2	ICS starting before, and <i>continuing through</i> the COVID period, straddling the pre- and during- COVID timescales.
Type 3	ICS emerging <i>as a result</i> of the new landscape, with impact arising because of COVID-19 (new or expediated).

Table 2: Summary of recommendations and relevance to combating risks to Impact

	Recommendation				
	Appropriate platform	Re-calibration and training	Increased role for Chair	Structured discussion times	Greater autonomy for smaller sub-groups
Temporal stability	x				
Development of shared values	x	x		x	
Prevention of outgroup silencing	x	x	x	x	x
Preventing the use of proxies and shortcuts		x	x	x	
Ability to assess COVID-effects			x	x	