Research Evaluation, 2025, **34(1)**, rvaf006
https://doi.org/10.1093/reseval/rvaf006
Article

Sceptics and champions: participant insights on the use of partial randomization to allocate research culture funding

Catherine Davies * **and Holly Ingram**

University of Leeds Research and Innovation Service, Nexus, Discovery Way, Leeds, LS2 3AA, United Kingdom

*Corresponding author. University of Leeds Research and Innovation Service, Nexus, Discovery Way, Leeds, LS2 3AA, United Kingdom.
E-mail: c.n.davies@leeds.ac.uk

Abstract

As part of the shift towards a more equitable research culture, funders are reconsidering traditional approaches to peer review. In doing so, they seek to minimize bias towards certain research ideas and researcher profiles, to ensure greater inclusion of disadvantaged groups, to improve review quality, to reduce burden, and to enable more transformative research. This paper presents the rationale and findings from a trial of partially randomized funding allocation (PRA) used to internally distribute Research England's Enhancing Research Culture fund at the University of Leeds. Of 26 eligible applications, six fell in the upper midfield on quality assessment and were randomized. Of this subset, one received the available funding. Qualitative data from applicants, reviewers, and moderators in the trial suggest modest gains regarding the reduction or perception of bias and efficiency of peer review. There were variable benefits of the feedback that PRA provides. A range of concerns emerged about PRA's perceived fairness, bluntness, and threat to quality. Based on the behavioural responses we captured, we present strategic and operational recommendations for professionals working in research evaluation who are interested in adopting PRA as a way of improving research culture. For example, we highlight the importance of (1) sharing data on the role of chance in traditional peer review, and (2) considering the benefits and risks of PRA at both group and individual applicant levels.

Keywords: grant peer review; research funding allocation; partial randomization; modified lottery; research culture.

1. Introduction

In the move towards a more equitable research culture, several funding bodies have started to reconsider traditional approaches to peer review. In doing so, they are seeking to minimize bias and ensure greater inclusion of disadvantaged groups, to improve review quality, to save time/reduce burden, and to enable more transformative research. UKRI's recent review of peer review (Kolarz et al. 2023) evaluates 38 intervention-types designed to optimize peer review processes. One of these approaches is partially randomized allocation (PRA) of funding, aiming to:

> [...] remove bias (both against demographic factors and riskier ideas), and to reduce administrative burden in the selection process. Mostly the burden is mentioned in connection to ranking, but the literature suggests that it has also been used (in connection with other interventions) to enable shorter applications.

The review justifies PRA thus:

> [...] increasingly overwhelming evidence that while peer/panel review reliably identifies the very highest quality applications, as well as the 'tail' of unsuitable low-quality ones, it tends towards arbitrary decision-making in the 'upper-midfield' of the quality spectrum.
>
> Kolarz et al. (2023: 44).

PRA's principles and implications raise several questions (Golberg 2022; Nature Editorial 2022; Harford 2023). For example, can we justifiably assess units on their grant income when the element of chance is explicitly introduced? Are human selection biases inherent in peer review merely shifted to the process of threshold-setting? Should applications that lost out due to randomization be allowed to resubmit? As innovation in research assessment expands as part of the growth in research culture initiatives, evaluative literature addressing these debates is emerging from researchers, meta-researchers, and professionals in research and innovation management and policy (Woods and Wilsdon 2021a,b; Kolarz et al. 2023).

As part of the research culture strategy at the University of Leeds, in 2023–4 we adopted PRA for our internal open call for research culture projects funded by Research England's Enhancing Research Culture fund. This meant that we effectively used a partial lottery system to make funding decisions for proposals considered equal on core quality criteria. Similar to other trials of randomization in research funding, we opted for its partial form, by which the middle tranche of applications rated on core quality criteria go through to random selection for funding.

The relatively small number of applications to our call enabled us to solicit opinions on the experience. To add to the growing body of investigations into PRA, we compiled our processes and emerging data to share the outcomes of our trial, to evaluate the impacts of the PRA pilot, and to make recommendations for future adaptations.

Contributory data includes indicative quantitative measures from the PRA round relative to the traditional peer review process used in our previous open call, as well as stakeholders' reflections on their experience of PRA.

1.1 Precedents for partially randomized allocation

Although random selection of projects for funding was discussed some time ago as part of RAND's publication on

alternatives to peer review (Guthrie at al. 2013), and exemplified by Brezis (2007) as 'focal randomization', its uptake to date has been limited. Kolarz et al.'s (2023) review highlights the use of PRA by a range of research funding bodies, including New Zealand's Health Research Council (see Liu et al. 2020 for a review), the Volkswagen Foundation in Germany, and the Swiss National Science Foundation. Among the first funders trialling in the UK are NERC (*Exploring the Frontiers* and *Pushing the Frontiers* schemes) and the British Academy (BA/Leverhulme Small Research Grants). NERC's stated aims were to increase the diversity of award recipients and streamline the application process, while the BA's rationale was to improve research culture through a transparent and simplified system, to remove human bias and partiality to achieve equity, to enable limited feedback to applicants, and to ease the burden on applicants and research officers without impacting the quality of applications and assessment. In summer 2023, Wellcome used PRA to allocate funding to applicants to its Institutional Fund for Research Culture (IFRC), to work towards removing biases and support fairness, 'recognising that there is no "right" answer for research culture but plenty of solutions' (Lewis-Wilson, Towers and Wykeham 2023).

PRA is also used in charity funding, e.g. Nesta's Explorations Initiatives. In their justification, Nesta describe funding as 'a nightmare of unconscious biases, popularity contests, conservative decision-making and trying to slide a cigarette paper between two applications that are equally as good […]'. Nesta states similar aims to the funders referenced above, i.e. to save time, reduce bias, improve diversity, and give space for more unconventional ideas.

PRA has had relatively low uptake, and little direct metaresearch with applicants and reviewers involved in PRA has been published to date [with the notable exception of Liu et al. (2020)]. However, some positive impacts are emerging. Kolarz et al. (2023: 44) report that at least two funders were found to have diversified their awardee pool. Applications to three more were found to increase in response to the introduction of PRA, reportedly due to a perceived higher chance of success among applicants. In Woods and Wilsdon's (2021a) small-scale qualitative study with six research funding bodies, the key driver for PRA was found to be fairness—both of decision-making and to applicants regardless of background or field. The study also revealed challenges regarding the communication of the system to stakeholders, with concomitant reputational risk.

In 2023, the British Academy (hereafter BA) published promising interim findings of its 3-year trial of PRA. From the two initial rounds (of six), successful candidates have come from a wider range of institutions, many in receipt of their first BA Small Grant, and more from Scotland and Northern Ireland than previously. There has been a notable increased in successful BAME applicants (27% vs. 18% under the previous system). Notably, the BA have seen a 32% increase in applications to the scheme, up from the preceding year when the traditional application process was used.

1.2 Our rationale for change

Our internal, open-call scheme for enhancing research culture funds has been running since 2022. The scope of the fund is broad and attracts a wide range of research questions, methodologies, activities, and team structures. This heterogeneity makes ranking more challenging than comparable calls with a narrower focus. Further, having observed the external trials in partially randomized allocation, we wanted to investigate the anticipated benefits for research culture more locally. Reflecting the BA's model, these benefits were:

1) *Reduction of bias*: Among strong applications that are deemed equally fundable, randomization should reduce conscious or unconscious bias against people, thus addressing inequities that may be experienced by e.g. early-career researchers or those from under-represented groups. Randomization should also reduce bias against particular research ideas, e.g. towards safer options at the expense of more radical proposals.
2) *Efficiency*: Partial randomization should ease the burden on reviewers as they need only to provide a simple but rigorous threshold judgement.
3) *Feedback*: Partial randomization provides basic feedback to unsuccessful applicants who do not pass the quality threshold, and to those who pass the threshold but lost out during the randomization process.

PRA was thought to be well-suited to our 2023–4 call based on the previous year's response of high quality, small-scale applications for exploratory studies typical of this scheme, together with the relatively small and time-poor review panel, and a tight reviewing timeframe.

The previous year's call used a more traditional reviewing process, requiring a panel of 20 reviewers to allocate 0–2 points against each of the 10 quality criteria, offer brief qualitative comments, then attend a full panel review session to come to final decisions.

2. Methods

This section describes the process by which we integrated PRA into the open call for projects.

2.1 Launch

We provided the three-part rationale for adopting PRA (as above) within the call guidance distributed to potential applicants.

2.2 Reviewing

We received 32 applications to the scheme. Fourteen peer reviewers were invited from the University's research culture governance groups and cohorts of the University's 100 Black Women Professors NOW programme. Reviewers were randomly paired and asked to reach agreement on approximately four applications each. Each pair was allocated a moderator from the research culture team to adjudicate disagreements (this support was only called on from one of the seven reviewer teams). Successful applications were required to pass an initial quality threshold: applications were deemed fundable if they satisfied the first two criteria plus a minimum of two others. This was determined via a light-touch peer review process involving six binary criteria:

1) Does the proposal persuasively articulate the research culture problem or challenge that it aims to address? (required)
2) Are its aims clear and achievable within the given timescale? (required)
3) Is the methodology appropriate?
4) Are the likely impacts of the project identified, and are the outcomes measurable?

5) Are the roles and responsibilities of all team members and any partners clearly defined?
6) Are the costs requested appropriate?

The threshold and criteria provided were designed to eliminate proposals which were out of scope, unclear, unfeasible, without impact, and/or poor value for money.

2.3 Randomization

As this was our first exploration of PRA, we decided to use a moderate form of randomization in which only the mid-graded applications are randomized (also used by Wellcome's IFRC), cf a more comprehensive approach in which all applications above a threshold are randomized, such as the form used by the BA.

Applications that received a 'yes' response from each reviewer pair on all six criteria clearly passed the quality threshold and were automatically funded.

Our protocol stated that if there were remaining funds after this allocation, proposals that received one 'no' response from the reviewer team on criteria 3–6 would be randomized. Those that came at the top of the randomized list down to the total funding limit would be recommended for funding. This process would be repeated for proposals that received two 'no' responses. Proposals receiving a 'no' on more than two criteria, or on criteria 1 and 2 at any stage would fail the quality threshold and be rejected.

NB. A subset of six applications were not randomized. These concerned projects that had been awarded pilot funding in the previous year's call, and which had then been submitted as follow-up applications. Five of these met the initial quality criteria and were funded without randomization.

2.4 Outcome

After the removal of the five successful follow-up applications, there were 26 remaining applications. Fourteen of these did not pass the initial quality threshold so were rejected.

Of the remaining 12, six passed all six quality criteria and were funded.

Six additional applications passed the quality threshold, passing five out of the six criteria. These were randomized as follows. A simple R script (R Core Team 2023) was written to generate a random list of numbers 1–6. In MS Excel, this list was pasted into a column next to the prepared list of project numbers, assigning a random number to each project number. The entire table was then sorted by the random number list in ascending order, generating a ranked order of priority for funding by project number.

The application at the top of the randomized list was funded. The funding allocation limit was then reached. Therefore, five applications were unsuccessful due to randomization. A total of £382K was allocated to 12 funded teams. One of these was awarded using PRA. Figure 1 shows the review and randomization process.

2.5 Diversity monitoring

Diversity information on each team member was requested on the application form, alongside the option to withhold this information. After randomization and funding recommendations, the research culture team manually collated and reviewed the gender, race, disability, and career stage diversity among project teams to verify that the PRA process had not generated any preferential biases among successful vs. unsuccessful applications (see Fig. 2). The outcome of diversity monitoring was presented to the reviewer panel. If there had

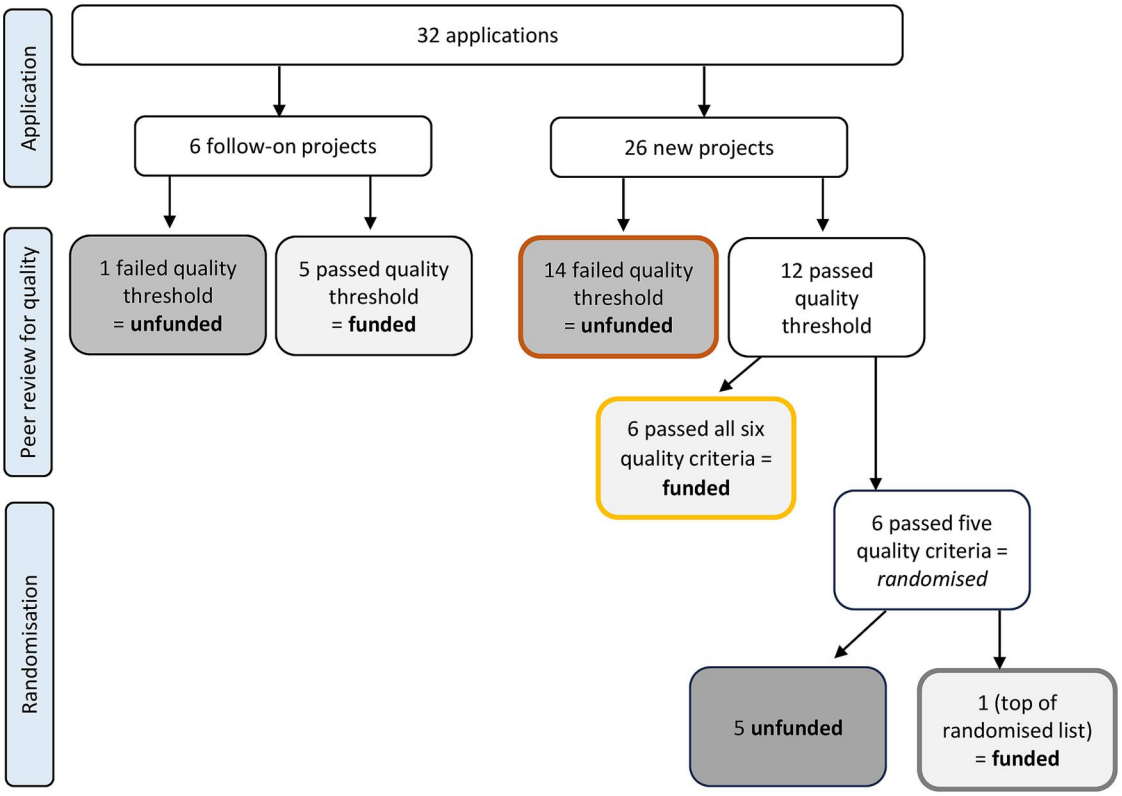

Figure 1. Flowchart of review and randomization processes.

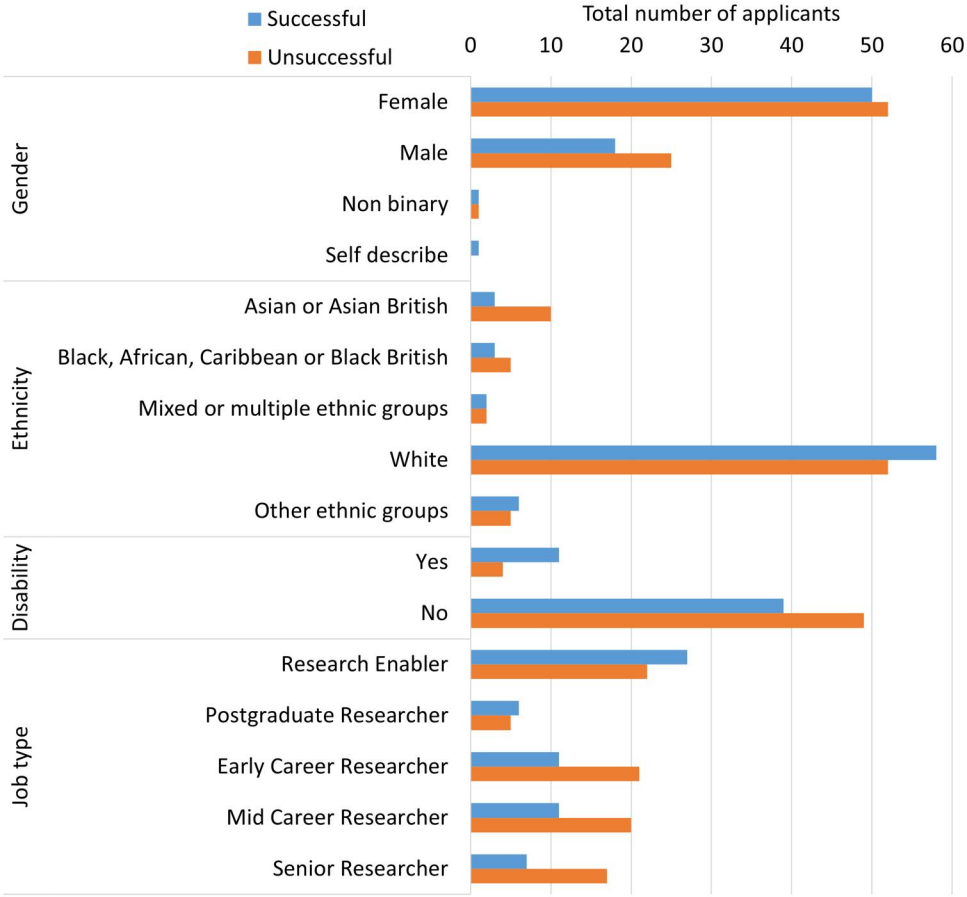

Figure 2. Diversity monitoring data. *Note*: Applicants provided diversity data on their applications in varying formats. This representation is as accurate as possible from the data provided.

been concerns regarding bias, i.e. if the randomization process had reproduced inequities towards a privileged group by chance, it would have been run and checked again. We included this exception in our protocol to protect against an unacceptable unintended consequence.

2.6 Notification

Applicants were notified of review outcomes and unsuccessful applicants received feedback. This indicated whether their proposals had been deemed fundable but not selected for funding due to randomization ('your proposal was deemed fundable but unfortunately came below the threshold of available funds'), or not selected for funding due to falling short of the quality threshold ('unfortunately on this occasion your proposal has not been selected for funding'). This was accompanied by written reviewer comments where provided.

3. Results

3.1 Stakeholder reception

As Table 1 shows, when comparing the 2022 (traditional) and 2023 (PRA) calls, a similar number of applications (35 vs. 32 respectively) and successful applications (13 vs. 12) were generated, though both the total amount requested and the mean cost per application was ~20% more in 2023, reflecting the higher value of the overall fund and better awareness of the scheme. The 2023 call required fewer reviewers and yielded significantly fewer incomplete reviews (0 vs. 15).

We elicited comments on the PRA process from applicants, reviewers, and moderators using a Microsoft Form questionnaire. See Supplementary Table 3 (Appendix) for the full list of questions used to elicit stakeholders' opinions on the PRA process. This was completed by reviewers and moderators during the reviewing panel session. Applicants received the questionnaire via email approximately a fortnight after the release of reviewing outcomes. Table 2 shows the frequency of responses from each stakeholder group, by role type and career stage.

Here we summarize the qualitative responses by question and theme, exemplified by selected data.

Did the PRA approach influence your decision to apply? If so, in what way?

The majority of applicants (7/12) reported no influence. Of those who did report an influence (3/12), this was negative, citing the explicit chance factor as a deterrent. Two respondents did not complete this section.

Yes, I almost did not apply as I was unsure if PRA was the right approach.

It made me slightly more nervous to apply and also a little frustrated that, technically, a proposal that would have scored less than ours using a standard scoring criteria could have been funded, and ours not.

On the whole, I found it a bit off-putting because it seemed that part of the process would be governed by chance rather than merit. I also didn't think the information on process

Table 1. Comparative quantitative data on *Enhancing Research Culture* fund processes.

	Round 1: December 2022 (traditional allocation)	Round 2: July 2023 (partially randomized allocation)
Number of applications submitted	35	32 (26 new; 6 follow-up)
£ value of applications	Total: £853,713 Range: £6K–£50K Mean: £24,392	Total: £1,053,000 Range: £10K–£50K Mean: £30,086
Number of applications randomized, i.e. passing quality threshold	N/A	6
No. applications awarded funding	13	12
Final success rate	13/35 = 37%	12/31 = 37%
Total fund available	£280,000	£400,000
Total funding awarded	£267,783 (99% of funding pot)	£382,000 (95% of funding pot)
Profile of successful/unsuccessful applicants (i.e. gender, race, disability, career stage, faculty)	Did not collect	See Table 2
Applicant team size	Smallest: 2 Largest: 14	Smallest: 2 Largest: 35
Range/consensus of reviewer scores	Large variance	Six out of seven reviewer pairs agreed on all scores before joint score submission. One pair called on the moderator to facilitate agreement on two applications.
Number of reviewers	20 total 3 per application (approx. 5 apps per reviewer)	14 total 2 per application (approx. 4 apps per reviewer)
Profile of reviewers (i.e. gender, race, disability, career stage, faculty)	Did not collect	2 senior members of staff 2 mid-career 3 did not specify 3 professional services 2 academic 1 technical 1 did not specify
Volume and type of queries from applicants	Approx 15 queries on: Application form, project type, buy out, costings, deadline, eligibility, signatures, future calls, application support	Approx 7 queries on: Future calls, external collaborators, project type, signatures, deadline extensions due to industrial action
Volume and type of queries from reviewers	3 queries on: Availability and scoring process	2 queries on: Binary judgements and HR timescales
Number of incomplete reviews	3 reviewers did not complete, amounting to 15 incomplete reviews	0

Table 2. Frequency of responses from each stakeholder group, by role type and career stage.

Number and type of stakeholders questioned	Number of responses	Breakdown of respondent role type	Breakdown of respondent career stage
31 applicants	12 (6 successful)	3 professional services 8 academic 1 did not specify	1 early career 4 mid-career 4 senior members of staff 3 did not specify
14 reviewers	7	3 professional services 2 academic 1 technical 1 did not specify	2 senior members of staff 2 mid-career 3 did not specify
5 moderators	6 (duplicate response by 1 moderator)	4 professional services 2 academic	3 senior members of staff 3 mid-career

provided beforehand was sufficient for me to understand what sort of PRA method was employed. We live in the age of AI—what model of randomization was used and how was this model trained? How biased might the model be?

Did the PRA approach affect the way you reviewed the applications? If so, in what way?

Although the reviewer response rate was low (50%), some said that PRA gave them a heightened sense of the importance of their decisions. They also found the binary rating unsatisfactory and lacking in nuance. Two reviewers reported no influence of PRA on their behaviour, one citing that they were '*still entrenched in the traditional approach*'.

- Increased awareness of power as a reviewer:

Ultimately it did not change anything but I was aware that the more applications that I scored as able to proceed to randomization, the less chance each application had of success.

It did make me think about the role I played as a reviewer and how my scores would affect chances of success. I think it impacted my scoring—when I was not 100% sure about my score, as I struggled to give either a 100% yes or no answer, I found myself hoping that some bids would still get a chance to be selected through the second stage random selection.

- Indirectly: the use of binary judgements was felt by some to lack nuance. This was mitigated by many reviewers by adding voluntary qualitative comments to their binary judgements (which they intended to be developmental for applicants).

It forced collective binary decisions which might have influenced the outcomes—if the reviewers had slightly different views there was a tendency to discuss and resolve to a single view.

No—but in combination with a binary approach I think it did, so I found myself wanting a third 'partially' option where it wasn't clearly a Y or a N as I was more aware of the consequences.

I find the binary scoring unhelpful and unsatisfying. It was the same for my co-reviewer. I would have felt more comfortable giving scores from 1–5; as in very rare cases was it as clear cut as a yes or no. That made it really hard to review as I feel the responsibility to give deserving applications a chance for success. I felt under more pressure to get the scoring right, but the only way to showcase my thinking was providing comments. I also think the comments are helpful for the recipients—both those successful and unsuccessful.

I found the binary score challenging in some cases and perhaps a 3 (or 4?) point score would have allowed more nuance.

Did the PRA approach affect the typical burden involved in reviewing? If so, in what way?

Reviewers' responses to this question went against our anticipated reduction in burden, with most (4/7) reviewers reporting that they spent the same amount of time and attention on the reviewers as they would using a traditional system. This was also reflected in the generous provision of optional comments in the reviews. The remaining three said they weren't sure as they didn't have enough experience to compare with the traditional reviewing system.

- *Probably not, as the panel meeting took time, but was actually really beneficial.*
- *No efficiencies garnered in terms of time and effort as only appropriate to review each in detail.*
- *It did not impact—still read the applications in detail and reflected on each criteria before responding to Y/N.*

Are you convinced by our rationale for trialling PRA, e.g. ***reduction of human bias, efficiency, provision of feedback, innovation of processes?***

The data suggest that stakeholder role was a factor in how convinced stakeholders were of the rationale for PRA. Applicants were the most sceptical (while also representing the full range of scepticism). Reviewers and moderators were more convinced, with the caveat that moderators may have been positively biased: they were all from the research culture team, who tend to champion PRA.

Applicants:

- Not at all convinced × 4 (2 successful applications; 2 eliminated on quality)
- Somewhat convinced × 3 (2 successful; 1 eliminated on quality)
- Mostly convinced × 2 (1 success; 1 n/a)
- Totally convinced × 2 (1 success; 1 eliminated on quality)
- Other × 1, commenting: *I don't think the rationale was explained as such—just presented as 'this is how we're doing it'*

Reviewers:

- Somewhat convinced × 1
- Mostly convinced × 2
- Totally convinced × 3
- Other × 1, commenting: *I can see the argument, but I am not convinced it worked or resulted in a fairer selection: the proposals I/we chose to go forward into the random selection pool, the ones I thought had greatest merit did not get selected, whilst those I felt were good enough but not as strong as my/our top choices got selected. This feels really dissatisfying and wrong.*

Moderators:

- Somewhat convinced × 1
- Mostly convinced × 2
- Totally convinced × 3

Do you have any concerns about the move to PRA?

This question elicited many comments. Some applicants reported being uncomfortable with the perceived element of chance, citing unfairness that careful work should be reduced to chance. Others felt that PRA was an overly blunt instrument, unable to merit certain ideas or applicants with particular profiles. The partial nature of randomization raised some concerns with some reviewers feeling uncomfortable that some applications (previous pilots and the highest scoring) did not undergo randomization. Two applicants were concerned that the binary scoring system may threaten quality (i.e. the bar for a 'yes' decision may be lower than a continuous scoring system). Another applicant felt that the feedback did not clarify the reasons for rejection. Eight respondents reported no concerns.

- 'Dumb luck'/lack of fairness

The reason why I put 'not at all convinced' is because of the pot luck of this funding allocation after the threshold has been met. I do think it is fantastic to trial innovative ways but when I saw this as a method on the guidelines it did feel me with anxiety. I am not sure whether it is fair to reduce the

amount of work and consideration that is taken on the proposals to a lottery chance of success. Could each application that is peer reviewed be given a grade and then you have a rank order in terms of quality and the top ones could then be discussed at a panel (like the AHRC peer review college). (applicant)

I've yet to see a convincing argument for PRA in any area. However much it is claimed to reduce bias and be more 'efficient' (which is a managerial weasel word that is blighting academia), the bottom line is that it makes life simpler for reviewers (abdicating difficult decisions about funding) and considerably more unfair for applicants. Funding bodies should be able to justify their decisions, not leave them to chance. (applicant)

If I had not been successful I am sure I would be very negative about the PRA. As it turns out, I was funded, but I am still unsure PRA is appropriate. (applicant)

Funding more on luck than merit—is it fair to cite these successes in promotions, etc? (moderator)

- Blunt instrument/inequity

PRA may not fully account for the diversity of projects proposed, and may not be able to strategically target important areas. (applicant)

Yes, while I appreciate the gesture towards equality in the move to PRA, and take your point about the drawbacks of human bias, I believe that we're increasingly as a University moving in the direction of equity, which is the (welcome) recognition that some come with greater disadvantage, and therefore may require case-by-case assessment which a PRA process is incapable of providing. Efficiency may come at the cost of equity in PRA. For example, there may be two applications of equal merit, but one which comes with more disadvantage factors/from a project team with protected characteristics. How would PRA account for this? In the same way that we're starting to recruit more equitably as a University, we should strive to [fund] projects more equitably as well. (applicant)

I agree that it will remove potential bias, unconscious and conscious, and this is very important. However, I think it could disadvantage earlier career researchers, who sometimes benefit from positive discrimination by reviewers, e.g. *they are given the benefit of the doubt more often to account for their relative inexperience.* (reviewer)

For PRA we need to understand how to pair reviewers up to assess applications: is two reviewers sufficient? Consider a third for each application. (reviewer)

A post-outcome review of diversity data is key to successful implementation: without it we risk inadvertently making things worse. (moderator).

- Concerns about the partial nature of randomization

I do wonder about moving to include the previously funded applicants and the top scorers in the randomization: it might be fairer and more cost effective if RE funding is reduced. I appreciate that we solicited pilots last year with the option of getting more but just not totally convinced we should fund any without the randomization. (moderator)

It is interesting that 6 of the applications were awarded 6/6 so went through without undergoing randomization. As the quality and number of applications increases we need to consider how we avoid rewarding those better at writing an application, for example an application may score 5/6 but be a really good idea not quite written as well as another less good idea written very well scoring 6/6

- Threat to quality

I totally get why the BA would do it for their small research grant scheme as it's so massively popular. PRA reduces the amount of time they have to spend deliberating on cases they really can't decide between and the metrics so far prove that it has encouraged more applications from previously under-represented groups. I'm less certain about its use for relatively low volumes of applications to decide between. I guess that if I can be reassured that the quality threshold is high enough then I think it's a good idea. (applicant)

The British Academy/Leverhulme format, of scoring proposals and then them going into a ballot for funding should they meet a certain score threshold, seemed a slightly better approach than a simple yes/no approach as used for this call. To me, it seemed quite easy to meet the 'yes' criteria, but I'm not sure this binary approach would reflect the standard of the proposals which meet the 'yes' criteria. (applicant)

- Lack of transparency:

The process wasn't very transparent. My outcome/feedback email was completely unclear whether our application had been part of the lottery or whether it didn't meet the threshold for consideration. This isn't particularly helpful. Information needs to be explicit and clear. (applicant)

- No concerns:

I think it's a fantastic and very fair approach (applicant)

I think it's a really good move! I've had so many funding applications rejected by large schemes without knowing why. It's helpful to know that a rejection might not be because of my application (or chosen co-applicants) but luck of the draw. (applicant)

No, it's the way forward (moderator)

I think the benefits are strong especially as this can work to reduce bias (moderator)

+ 2 reviewers (no comments)
+ 2 moderators (no comments)

Do you have any suggestions to help us improve our system of PRA?

Most comments reflected the need for greater transparency of feedback, a return to full peer review panels on equity grounds, a more nuanced scoring range, or a right to reply.

Be clear about whether an application failed at the quality threshold or at randomization (applicant)

I'm not yet convinced that PRA is a sound replacement for human processes, and would be in favour of either a return to full peer review with unconscious bias observers or the introduction of some means of equity monitoring at the PRA stage of the process. This would of course confound the

'random' part of the process, but would be more equitable. (applicant)

Ditch it. But if you are going to make life easier for reviewers and things more 'efficient', then the least that could happen is that the feedback could be more constructive and clear on why the reviewers have made their recommendations. At present, the system seems to work against applicants in both the transparency of selection (if recommended for funding; obviously not an issue for my team!) but also in terms of the quality of feedback. (applicant)

If there are projects that are borderline above/below the threshold perhaps a peer group panel makes the final decision. (applicant)

Have an opportunity to respond to reviewer comments before deciding which proposals meet the quality threshold. (applicant)

I would apply a different, non-binary scoring system; I would also not rely on two criteria only to decide whether the bids make it through to the next round but all of the criteria; I would also introduce a criterion on return in investment as I think this is missing. (reviewer)

Thought needs to be given to the reapplication/resubmission policy for those proposals that passed a given quality threshold but have missed out on funding due to randomization—are they permitted to resubmit? Where is the cut off? (reviewer)

Need to have some contingency plan on if more applications pass the quality review than you have funding. In this case would we revert to full randomization? (moderator)

4. Discussion

The anticipated benefits of PRA stated in our rationale were reduction of bias; efficiency; and feedback. Here we reflect on the extent to which our trial achieved these.

4.1 Reduction of bias

As randomization was only used for applications passing five of the six quality criteria, this reduction could only be partial. That is, reviewer bias for certain topics could have influenced initial binary decisions, leaving all those with six 'yes' responses with a straightforward route to funding. Bias should have been reduced for applications scoring at the next level down, though this only concerned six of the 25 applications. Thus, any reduction of bias was modest.

A system in which the most highly rated applications receive funding (i.e. partial randomization) is inevitably subject to a degree of reviewer bias towards certain ideas. On balance, this is preferable to a fully randomized or lottery system which does not use a quality threshold.

Because qualitative comments were not based on detailed knowledge of outcomes, they could not conclusively address the question of bias. However, some comments revealed concerns about the removal of positive bias towards disadvantaged groups, e.g. ECRs or researchers of colour. However, it is not clear to us how such positive bias (or 'benefit of the doubt' as one respondent noted) might play out under non-randomized, anonymized systems, outside of positive action or quota schemes.

Diversity monitoring indicated that randomization had not generated any notable bias according to gender, ethnicity, disability, or job type, and demonstrated an even distribution of unsuccessful and successful applicants across the areas listed (see Fig. 2). Due to other aspects of the call management differing from the traditional approach in the previous year, it is not possible to compare diversity outcomes. We hope to return to this comparison in future rounds of the scheme.

4.2 Efficiency

Surprisingly, most reviewers reported that they spent the same amount of time and attention on their reviews as they would using a traditional system. This finding contrasts with the aim to reduce reviewer burden, stated by this study and in other trials of PRA (Kolarz et al. 2023: 44). PRA's intended reduction of burden makes intuitive sense, since reviewers don't need to stratify similarly strong applications. However, on reviewing the extant literature, we find little evidence for efficiency. Although respondents comment on the *potential* for efficiency, actual benefits seem negligible. In some cases, PRA was *more* burdensome due to its novelty (Woods and Wilsdon, 2021b: 17–18). It will be important for future trials to scrutinize this anticipated benefit with cohorts going through PRA. Regarding the current study, the lack of relief may be related to both the small-scale and the internal nature of the funds: institutional colleagues were keen to provide developmental feedback. If this finding is replicated more widely, it presents implications for the case for PRA. It also nullifies some of the scepticism about the rationale for PRA, i.e. it being solely to reduce reviewer burden.

Another important way to view burden is that placed on the administrative team, who required fewer reviewers for the PRA round, and who found it easier to recruit those reviewers, and who were more compliant in returning their reviews. Thus, the overall burden was lighter, even if the burden on the reviewers who took part was similar.

4.3 Provision of feedback

Although we provided basic feedback on reasons for rejection (randomization vs. quality, wording provided in section 2.6), some applicants found this confusing. On reflection, our notifications should have (1) been clearer that 'not selected for funding' meant that these applications had fallen short of the quality threshold, and (2) provided reviewer decisions relating to each of the six quality criteria to enable focused revisions in any future applications. Administrators would need to be clear about applicants' right to respond in this scenario.

Wider discussion with colleagues also suggested that being rejected on the grounds of randomization is both (1) easier to accept and (2) frustrating in that no improvements can be made when resubmitting.

In addition to the hypothesized advantages of PRA (i.e. reduction of bias, increased efficiency, and increased feedback), an unanticipated theme came through strongly in our data: the perceived element of chance and associated feelings of unfairness. Although randomization seems to bring this theme to the fore, it is important to highlight the arbitrariness of traditional peer review process, incorporating a degree of uncertainty by reviewers. Evidence from higher-volume calls demonstrates that inter-rater reliability between reviewers and panels is low especially at the funding threshold (Heyard et al. 2022), and the ranking of close-scoring applications is scarcely better than chance (Fang, Bowen and Casadevall 2016; Fang and Casadevall 2016; Jerrim and Vries 2020). With this in mind, PRA does not seem to diverge from how

traditional processes treat the strongest applications (Golberg 2022). In line with several funding agencies now increasingly acknowledging the role of chance (Heyard et al. 2022), it will be important to make the somewhat 'random' or chance nature of traditional peer review clear to stakeholders when considering potential benefits and risks of PRA (as well as the challenge of ranking heterogenous proposals). This quote from Stafford et al. (2023) summarizes the gains that PRA can bring to fairness of the process:

> A recent study (Woods and Wilsdon 2021b) found that the strongest motivator for funding institutions to use partial randomisation is fairness: a fairer decision-making process when peer review had reached its limits; fairer to applicants, as it is blind to institution, geographical location, race, gender, discipline and methodology; and also a transparent process and therefore easier to communicate and understand funding decisions. Other organisational motivators are the desire to break deadlocks in, or reduce time spent on panel decision making, and to ameliorate risk aversion or other concentrations of awards so as to facilitate the funding of a greater plurality of research topics and methodological types.

The chance factor also throws up concerns about reputational impact from both rejections and successes in PRA. Some applicants may worry about crediting their awards (if successful) to randomization, or being judged negatively if they lost out due to randomization. Where traditional academic CVs are still required, a brief line to clarify when randomization was used in a grant scheme may help to mitigate this concern.

The qualitative data suggests that trialling PRA had no effect on the volume of applications submitted, though it was correlated with a smoother reviewing process (note that the 2023 round used pair-reviewing which may have been a stronger causal factor in reviewer compliance).

Overall, our sample was less enamoured with PRA than some of the literature would predict. PRA has been heralded as a means of addressing deeply entrenched, problematic aspects of the funding system, e.g. bias; unreliability; conservatism (Brezis 2007; Fang and Casadevall 2016; Golberg 2022; Heyard et al. 2022; Nature Editorial 2022). The current study demonstrates the need to go beyond 'what-if' discussion (Barlösius and Philipps 2022; Philipps, 2022), and instead engage meaningfully with applicants experienced in PRA, e.g. Liu et al. (2020), who found their cohorts to accept the change, albeit conditionally.

5. Conclusions

Based on questionnaire data responding to our trial of PRA, only modest gains were made regarding the **reduction of bias.** NB. there was limited potential to achieve this in our trial as only one application was ultimately funded due to randomization. An alleged loss of positive bias towards marginalized groups was seen as a drawback (although this is hard to analyse since positive action was beyond the scope of the trial). Although reviewer burden did not differ from traditional reviewing, the PRA round was **more efficient** to run by the administrative team. The benefits of the **feedback** that PRA affords varied significantly among applicants, so are inconclusive.

Behavioural responses to changes in funding mechanisms are important to understand. Given that only 6/32 applications were randomized and available funds only covered one of those six (since the application at the top of the randomized list was high-budget), it is important to interpret our findings with an awareness that the documented PRA effects are driven by **perceptions** of applicants and reviewers, rather than the randomization process itself.[1]

Our trial was based on immediate experiences of the **allocation process** rather than the longer-term impact of PRA on the nature and outcomes of the funded research. Downstream effects on reduction of bias, gains in efficiency, or trust in funding allocation should be monitored. Randomization as a factor in funding success will be added to the routine evaluation of the outcomes and impacts of the funded project.

It is important to consider findings of PRA evaluations on two levels. At the group level, positive effects are emerging, e.g. The BA's diversification of successful applications. However, at the individual level, the process can be frustrating, e.g. perceptions of powerlessness by applicants.

This case study includes several limitations. First, it analyses effects only on stakeholders rather than on the wider research or funding system. These effects are also very early-stage. Second, the case study does not purport to be a controlled trial comparing PRA to the traditional peer review process used in our previous call: the two calls differed in several other aspects, e.g. reviewing criteria, concurrent industrial action in 2023, and a different applicant cohort. Third, relating to the stated aim to reduce bias, some reviewers may have been able to identify applicants despite anonymization, due to belonging to the same University community (though note that any conflicts of interest on the part of reviewers were expressed before the review process).

6. Recommendations

We will carefully consider the value of PRA in future funding schemes based on our findings in this trial. We note several recommendations for ourselves and our partners interested in trialling PRA.

6.1 Strategic

1) Consider exemption from randomization for marginalized groups to restore equity.
2) Reconsider exempting follow-on proposals from randomization due to questions of fairness.
3) In future rounds of PRA for research culture projects, communicate with stakeholders the evidence demonstrating the influence of chance in traditional peer review, as well as the specific challenges of assessing heterogeneous projects.
4) Consider the benefits and risks of PRA at both group and individual applicant levels.
5) In future rounds that use the consensus approach between reviewer pairs, capture information about their level of agreement as a measure of uncertainty. This can then inform decisions on reasonable levels of randomization, e.g. consider exempting from randomization applications with lower reported levels of agreement.

6) Solicit longer-range views on the outcomes of PRA from successful applicant teams.
7) Re-run the trial in the subsequent round of this scheme to investigate whether the marginal gains this year are replicated.
8) Share the findings of this case study with colleagues within and beyond the institution to facilitate discussion about the merits of PRA for different funding schemes, and to encourage consideration of alternative approaches to peer review.

6.2 Operational

1) Clearly include the reasons for rejection (i.e. randomization or quality, plus reviewer decisions on criteria) in outcome notifications.
2) Reconsider binary criteria, which some reviewers found more challenging to use than graded scoring (while for others it reduced the work required greatly). The granularity of the scale should be considered alongside the volume of applications: a finer-graded scale may be warranted in higher-volume calls to sufficiently discriminate quality.
3) Ensure consistency in the way that diversity monitoring data is provided by teams.

Acknowledgements

We would like to thank all the applicants, reviewers, and moderators involved in the 2023–4 open call for research culture projects at the University of Leeds who generously shared their experiences for this study. We acknowledge the research culture team at the University of Leeds for supporting the production of this report, and attendees at a meeting discussing partial randomization in research funding allocation, hosted at The British Academy in December 2024. Thanks to Helen Pickard (Research and Innovation Service, University of Leeds) and Steven Wooding (Research Strategy Office, University of Cambridge) who provided peer feedback on an earlier draft. Thanks also go to the research community at the University of Leeds who are developing unique areas of research culture practice to make meaningful, lasting change.

Supplementary data

Supplementary data are available at *Research Evaluation Journal* online.

Author contributions

Catherine Davies (Conceptualization, Methodology, Supervision, Writing—Original Draft, Writing—Review & Editing), and Holly Ingram (Investigation, Data curation, Methodology, Project administration, Writing—Review & Editing).

Funding

This work was supported by Research England.

Note

1. We thank an anonymous reviewer for highlighting this aspect of the study.

References

Barlösius, E., and Philipps, A. (2022) 'Random Grant Allocation from the Researchers' Perspective: Introducing the Distinction into Legitimate and Illegitimate Problems in Bourdieu's Field Theory', *Social Science Information*, 61: 154–78. https://doi.org/10.1177/05390184221076627

Brezis, E. S. (2007) 'Focal Randomization: An Optimal Mechanism for the Evaluation of R&D Projects', *Science and Public Policy*, 34: 691–8.

Fang, F. C., Bowen, A., and Casadevall, A. (2016) 'NIH Peer Review Percentile Scores Are Poorly Predictive of Grant Productivity', *ELife*, 5: 1–6. https://doi.org/10.7554/eLife.13323

Fang, F. C., and Casadevall, A. (2016) 'Research Funding: The Case for a Modified Lottery', *mBio*, 7: e00422–16. https://doi.org/10.1128/mBio.00422-16

Golberg, A. (2022, November 15) *The (Partial) Rise of (Partial) Randomisation*. Research Professional News. https://www.researchprofessionalnews.com/rr-news-uk-views-of-the-uk-2022-11-the-partial-rise-of-partial-randomisation/

Guthrie, S. et al. (2013) *Alternatives to Peer Review in Research Project Funding: 2013 Update*. Santa Monica, CA: RAND Corporation. https://www.rand.org/pubs/research_reports/RR139.html

Harford, T. (2023, September 29) *Sometimes, a Random Solution is Best*. Financial Times. https://www.ft.com/content/61f67834-c228-44bb-b1e9-9a60d6b57d27

Heyard, R. et al. (2022) 'Rethinking the Funding Line at the Swiss National Science Foundation: Bayesian Ranking and Lottery', *Statistics and Public Policy*, 9: 110–21. https://doi.org/10.1080/2330443X.2022.2086190

Jerrim, J., and Vries, R. (2020) 'Are Peer-Reviews of Grant Proposals Reliable? An Analysis of Economic and Social Research Council (ESRC) Funding Applications', *The Social Science Journal*, 60: 91–109. https://doi.org/10.1080/03623319.2020.1728506

Kolarz, P. et al. (2023, July 3) *Review of Peer Review*. UKRI. https://policycommons.net/artifacts/4385753/ukri-050723-reviewofpeerreview/5182293/

Lewis-Wilson, S., Towers, S., and Wykeham, H. (2023) 'The Luck of the Draw: Wellcome's Institutional Fund for Research Culture' [Version 2; Peer Review: 1 Approved, 2 Approved with Reservations]', *Wellcome Open Research*, 8: 525. https://doi.org/10.12688/wellcomeopenres.20057.2

Liu, M. et al. (2020) 'The Acceptability of Using a Lottery to Allocate Research Funding: A Survey of Applicants', *Research Integrity and Peer Review*, 5: 3. https://doi.org/10.1186/s41073-019-0089-z

Nature Editorial (2022, September 20) 'The Case for Lotteries as a Tiebreaker of Quality in Research Funding', *Nature*, 609: 653. https://doi.org/10.1038/d41586-022-02959-3

Philipps, A. (2022) 'Research Funding Randomly Allocated? A Survey of Scientists' Views on Peer Review and Lottery', *Science and Public Policy*, 49: 365–77. https://doi.org/10.1093/scipol/scab084

R Core Team (2023) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Stafford, T. et al. (2023) 'Where Next for Partial Randomisation of Research Funding? The Feasibility of RCTs and Alternatives', *Wellcome Open Research*, 8: 309. https://doi.org/10.12688/wellcomeopenres.19565.1

Research Evaluation, 2025, 34, 1–10
https://doi.org/10.1093/reseval/rvaf006
Article