


RESEARCH

Open Access



# Grant reviewer perceptions of the quality, effectiveness, and influence of panel discussion

Stephen A. Gallo<sup>1\*</sup> , Karen B. Schmalings<sup>2</sup>, Lisa A. Thompson<sup>1</sup> and Scott R. Glisson<sup>1</sup>

## Abstract

**Background:** Funding agencies have long used panel discussion in the peer review of research grant proposals as a way to utilize a set of expertise and perspectives in making funding decisions. Little research has examined the quality of panel discussions and how effectively they are facilitated.

**Methods:** Here, we present a mixed-method analysis of data from a survey of reviewers focused on their perceptions of the quality, effectiveness, and influence of panel discussion from their last peer review experience.

**Results:** Reviewers indicated that panel discussions were viewed favorably in terms of participation, clarifying differing opinions, informing unassigned reviewers, and chair facilitation. However, some reviewers mentioned issues with panel discussions, including an uneven focus, limited participation from unassigned reviewers, and short discussion times. Most reviewers felt the discussions affected the review outcome, helped in choosing the best science, and were generally fair and balanced. However, those who felt the discussion did not affect the outcome were also more likely to evaluate panel communication negatively, and several reviewers mentioned potential sources of bias related to the discussion. While respondents strongly acknowledged the importance of the chair in ensuring appropriate facilitation of the discussion to influence scoring and to limit the influence of potential sources of bias from the discussion on scoring, nearly a third of respondents did not find the chair of their most recent panel to have performed these roles effectively.

**Conclusions:** It is likely that improving chair training in the management of discussion as well as creating review procedures that are informed by the science of leadership and team communication would improve review processes and proposal review reliability.

**Keywords:** Bias, Communication, Discussion, Grant applications, Innovation, Leadership, Peer review, Research funding, Survey, Team science

## Background

The US National Institutes of Health (NIH) utilizes a “long standing and time-tested system of peer review to identify the most promising biomedical research [[1], p. 2],” as do many major research funders [2]. However, many reports of poor inter-rater reliability suggest a high

degree of subjectivity to the process [3–6]. One common procedure used to mitigate this subjectivity is to discuss each proposal at a meeting of the entire review panel, utilizing a larger set of expertise and perspective than just that of the 2–3 reviewers that are explicitly assigned to read and evaluate the proposal pre-meeting [7, 8]. In this system, final scores are typically generated from the average post-discussion scores of all assigned and unassigned reviewers (without conflicts of interest), in an

\* Correspondence:

<sup>1</sup>Scientific Peer Advisory and Review Services, American Institute of Biological Sciences, Herndon, VA, USA

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

attempt to ensure all available expertise is brought to bear on the final evaluation of the proposal [1].

Despite this intended goal, several studies have reported that discussion can have a somewhat limited effect on the final scoring of proposals [5, 8–11], with some studies estimating that a proposal's funding status (score above or below the funding line) is shifted from pre- to post-discussion for only 10–13% of proposals [5, 9]. While some studies [12] suggest that most discussions do yield changes in assigned reviewers' scores (which in general move closer together, narrowing the range of scores), the magnitude of scoring shifts after discussion are typically relatively small and are even smaller (with shorter discussion times) in teleconference panels compared to face-to-face panels [9, 13]. It is known that face-to-face communication is a richer channel than other virtual alternatives [14], and it is likely that the quality of panel communication plays an important role in how influential the panel discussions are on scoring.

Interestingly, it has been observed that scoring shifts are correlated with instances of panel discourse “in which reviewers made explicit references to the scoring habits of fellow panelists,” which have been referred to as score calibration talk [[8], p. 11]. These scoring shifts are particularly related to discussion where reviewers were being held accountable by other reviewers for how they calibrated their score relative to the descriptors in their written critiques. For instance, “your comments are meaner than your score” [[8], p. 11]. However, it is not clear how well these types of interactions are promoted by the chair or if all reviewers (especially unassigned reviewers) feel they have the opportunity to present such opinions. In fact, others have noted that, although the goal of convening panels is to bring a range of expertise to bear on the evaluation of a research proposal, opportunities are limited for dialogues between reviewers of different expertise [15].

In addition, no studies have reported if reviewers think the panel discussion sufficiently focuses on differences in opinion between assigned reviewers, so that unassigned reviewers are well-informed in their final scoring decisions. While typically unassigned reviewers' scores closely mirror assigned reviewers' scores [9, 10], some assigned reviewers exert more influence than others, which may be due to “expertise, authority and debating, persuasion, or argumentation skills” [[5], p. 51]. The chair's role is to facilitate discussion, moderate individual personalities, and provide a fair and balanced presentation of each proposal under evaluation to unassigned reviewers. However, it is not clear if all chairs effectively fulfill these responsibilities, and studies have not been undertaken to evaluate the ability of panel chairs to

steward the peer review panel proceedings. Such evaluations are, therefore, needed to better understand review panel facilitation, particularly since this type of information cannot be derived from peer review scores.

Little data are available regarding reviewers' perceptions of panel discussions. Some favorable reviewer perceptions have, however, been reported by NIH, specifically that 81% agreed or strongly agreed that “scientific discussions supported the ability of the panel to evaluate the applications being reviewed” [[16], p. 2], but the general nature of this statement does not provide a finer-grained understanding of panel discussions. Details of interest include the level of participation in discussion, whether discussion helps to clarify opinions, how well the chair facilitates this discussion, and how well unassigned reviewers can make informed decisions based on the discussion. Moreover, it would be pertinent to determine if reviewers feel that the discussions affect the outcome of proposal reviews.

To address these gaps, we developed a survey focused on reviewer perceptions of their most recent panel meeting experience and distributed it to a diverse group of research scientists. Two publications have resulted from analyses of the survey responses [17, 18], but neither publication focused on the section of the survey that addressed perceptions of the quality and facilitation of panel discussion and their impact on review outcomes. To examine these topics, feedback from the surveyed scientists were summarized regarding the quality, effectiveness, and influence of their most recent panel discussions with the goal of developing a better understanding of reviewer perceptions of panel facilitation to help inform the future implementation of review formats and procedures.

## Methods

### Survey

This study involved a diverse group of biomedical research scientists who responded to a survey. The survey was reviewed by the Washington State University Office of Research Assurances (Assurance# FWA00002946) and granted an exemption from IRB review consistent with 45 CFR 46.101(b)(2). Participants were free to choose whether or not to participate in the survey and consented by their participation. They were fully informed at the beginning of the survey as to the purpose for this research, how we acquired their email address, and the importance and intended use of the data. The general survey methodology has been described in two other manuscripts [17, 18]. The original survey contained 60 questions and was divided into 5 subsections; data from only 3 sections are presented in this manuscript to address the quality of peer review panel discussion: (1) grant submission and peer review experience;

(2) reviewer attitudes toward grant review; and (3) peer review panel meeting proceedings. The questions regarding discussion effectiveness, quality, and influence included here were not analyzed in the previous publications, although other aspects, such as review frequency and reviewer preference, were reported previously. Discussion effectiveness was defined as allowing for reviewer participation (both assigned and unassigned reviewers); clarification of the original opinions of the assigned reviewers; and the whole panel (assigned and unassigned reviewers) to be well-informed before scoring a proposal. Discussion quality was defined by the levels of openness and balance in the discussion, lack of bias, and focus and efficiency on key points. Discussion effectiveness and quality were of interest in general and specifically in relation to facilitation by the chair. The influence of the discussion was defined as impacting the final outcomes and promoting the best science forward.

The survey questions related to these definitions had either nominal (yes/no) or ordinal (Likert rating) response choices. For example, on a scale of 1–5 (1 = most definitely, 5 = not at all), did the grant application discussions promote the best science? However, respondents were also given the choice to select no answer/prefer not to answer. At the end of each section, respondents could clarify their answers in a free form text box. A full copy of the peer review survey is available in the **S1 File in the Supporting Information**. The raw, anonymized data are also available (<https://doi.org/10.6084/m9.figshare.8132453.v1>).

As described in previous publications, the survey was sent out in September of 2016 to 13,091 individual scientists from the American Institute of Biological Science's (AIBS's) database through the use of Limesurvey (© Hamburg, Germany), which de-identified the responses from respondents. AIBS's database has been developed over several years to help AIBS recruit potential reviewers for evaluation of biomedical research applications for a variety of funding agencies, research institutes and non-profit research funders. Most of these reviews are non-recurring and scientists are recruited based on matching expertise to the topic areas of the applications. The individuals invited to this survey were reviewers for AIBS (26%), had submitted an application as a PI that was reviewed by AIBS (62%), or both (12%). The 13,091 invited individuals represent the population who met the criteria of reviewer or applicant. Depending on the question, respondents were asked to focus on either the most recent peer review or reviews that occurred in the last 3 years.

#### Procedure and data summarization

The survey was open for two months: the initial invitation was sent on September 7, 2016, and a reminder was sent a month and a half later (October 24/25, 2016), and

the survey was closed 2 weeks later on November 7, 2016. Responses were then exported and analyzed using Stat Plus software. For this paper, participants were included only if they completed the survey and included an answer for questions 2e and 2f, which focused on whether they had participated in a peer review panel in the last 3 years and, if so, how often. Thus, all questions included in this analysis were focused on reviewer experiences. Reviewers were asked questions related to the qualities of panel discussion. Median and percentage comparisons were analyzed using non-parametric tests (e.g., Mann-Whitney, chi-square tests), due to the highly skewed ordinal distributions (most are > 1.0). Standard 95% confidence intervals (CI) were calculated for the Likert responses (for proportion data, binomial proportion confidence intervals were calculated). Effect size ( $d$ ) was calculated via standardized mean difference for all comparisons. Differences between groups were considered significant if there was either no overlap in CI or if there was overlap yet a test for difference indicated a significant result ( $p < 0.01$ ).

All comments made by respondents from the survey related to the Peer Review Panel Meeting Proceedings section were extracted. All quotes were then grouped according to which of the six specific questions analyzed in this manuscript they most closely addressed (Table 1). Multiple groupings per quote were permissible. Many of the quotes in this section were not relevant to the six questions or the survey section (e.g., their level of participation) and were therefore not considered for inclusion in this analysis. The remaining quotes were then examined for common themes, and when multiple quotes related to a similar idea were identified, quotes were selected for inclusion in the manuscript that were well-written, clear and specific relative to the survey question. If contrasting views on the same theme were expressed, care was taken to ensure that both quotes were included in the manuscript. If only part of a respondent's comment was relevant to one of the six questions, only the relevant portion was included. One author conducted the initial grouping of comments, but consensus among all authors was achieved before finalization in the manuscript.

Sorting of survey comments according to relevance to major survey questions considered in this analysis (total  $N = 153$ ). Results are displayed as  $N$  comments relevant to a specific question and % of total comments. Multiple groupings per quote were permissible. Comments could have positive components, negative components, both or neither (neutral). The proportions of total comments per question with positive, negative, or neutral sentiments are listed

## Results

### Response rate, demographics, and post-reminder response

Of the 13,091 individuals contacted for this survey, 1231 responded, giving a 9.4% response rate. Of the 1231

**Table 1** Response sorting

Question	% of total (N)	% Positive (N)	% Negative (N)	% Neutral (N)
Q1. Did the grant application discussions facilitate reviewer participation?	6.5% (10)	50% (5)	60% (6)	0% (0)
Q2. How useful were the grant application discussions in clarifying differing reviewer opinions?	7.2% (11)	64% (7)	55% (6)	9% (1)
Q3. Was the format and duration of the grant application discussions sufficient to allow the non-assigned reviewers to cast well informed merit scores?	15% (23)	52% (12)	65% (15)	0% (0)
Q4. How useful was the Chair in facilitating the application discussions?	24% (39)	59% (23)	51% (20)	15% (6)
Q5. Did the grant application discussions affect the outcome?	11% (17)	35% (6)	76% (13)	12% (2)
Q6. Did the grant application discussions promote the best science?	12% (19)	11% (2)	95% (18)	5% (1)
None of the above	41% (63)	8% (5)	48% (30)	46% (29)

respondents, only 874 of these completed questions 2e and 2f, 671 (77%) of whom indicated they had recently reviewed on a panel in the last 3 years. These 671 reviewer respondents formed the core group upon which the current analyses are based. Of these, free text answers were provided by 153 respondents, which were grouped according to relevant question and positive, negative, or neutral sentiment (Table 1). A total of 29 quotes were used in relation to specific survey questions; these are presented below. It was observed that respondents' comments related to questions 5 and 6 were overwhelmingly negative and, for those who made a comment, their median responses to Q5 (3.0, 95%CI 2.4 to 3.6) and Q6 (3.0, 95%CI 2.5 to 3.5) were more negative than those who did not make a comment for Q5 (2.0, 95%CI 1.9 to 2.1), and Q6 (2.0, 95% CI 1.9 to 2.1), respectively.

Demographics were analyzed in detail in previous publications: respondents were 66% male, 80% PhD, and 69% in a late career stage (e.g., tenured full and emeritus professorship) with the majority being age 50 or older (75%; median age 55), Caucasian (76%), and working in academia (81%). These respondents participated in an average of  $4.0 \pm 0.08$  panel meetings over the last 3 years.

The results from early respondents to the survey from the first month and a half were compared to results from participants that responded in the 2 weeks after the reminder email was sent out (the early respondent group comprised 60% of the final respondents). Median and percentile values for all six questions were nearly identical between the early and late respondent groups (Supplemental Table 1).

### Discussion effectiveness and quality

Utilizing the expertise of the whole panel is one of the rationales for discussing proposals at the meetings. As one respondent put it:

“Discussions can occasionally develop a herd mentality, so having multiple perspectives represented and facilitation to hear all voices is crucial.”

The vast majority (92%, 95%CI 90% to 95%) of reviewers felt the panel discussions involved reviewer participation (Q1), although participation may be in reference to the engagement of assigned reviewers. Some respondents specifically mentioned participation from unassigned reviewers, although their participation in discussion was not always at a high frequency:

“Overall, discussion was driven by the assigned reviewers. Sometimes an interested non-assigned reviewer would get involved. Sometimes, non-assigned reviewers would ask clarifying questions.”

“Non-assigned reviewers rarely ask questions or comment.”

Seventy percent of all reviewers felt discussions were mostly useful or very useful in clarifying opinions (median is 2.0, 95%CI 1.9 to 2.1 on a scale of 1 to 5; Q2). Several respondents remarked on the clarity of presentations and their importance in the evaluation:

“It makes a big difference to hear what the reviewers that carefully examined the application thought and why when there's some disagreement. People see different things. Bringing those things out in discussion helps assure more fairness of scores across applications when different people are reviewing.”

However, some respondents felt the distribution of comments from panel members and their relative weight on panel opinion was not even across reviewers and sometimes dependent on whether the assessment was positive or negative:

“My experience has been that folks do not use the discussion time most effectively. The primary reviewer is accepted at face value and when a dissenting opinion is voiced, the panel seems to be reluctant to discuss and rather defers to the primary reviewer.”

This is a flaw that makes the review only as good as the thoroughness of ONE reviewer. And in my limited experience I have seen some shoddy reviewing. This is unfair to the team that prepares the application.”

“On the whole, I felt most panel members were too polite and unwilling to offer frank opinions of weak proposals.”

As above, respondents commented on the usefulness of the opinions and interactions of unassigned reviewers in panel discussion:

“It’s always difficult to strike a balance between having non-assigned reviewers contribute to a discussion and their relative lack of knowledge of the area. However, my opinion is that they often help to clarify points that (maybe) are obvious to the expert reviewers but probably not to the rest of the panel, resulting in a more informed final score.”

Moreover, most reviewers (79%, 95%CI 75% to 82%) agreed that the format and duration of the grant application discussions was sufficient to allow the non-assigned reviewers to cast well informed merit scores (Q3). However, some respondents suggested that these discussions are only effective in influencing unassigned reviewers’ scoring:

“Discussions change very few minds among reviewers and probably are only useful to informing other, non-assigned panel members.”

“While the discussion does not alter the assigned reviewers initial scores, it provides context and rationale for the scores which helps the other reviewers decide on a score.”

Others lament that unassigned reviewers do not read the grants, and therefore their scoring is superficially based on limited discussion, potentially leading to bias:

“There really is not enough time for non-assigned reviewers to be able to read the grants and listen to the discussion. I think people just decide which of the assigned reviewers they like more (or whose argument they like more) and then vote with them.”

“It is not clear that the content of assigned reviewer comments are driving the scoring decisions of non-assigned reviewers”

“Very difficult for non-assigned panel members to actually judge the application fairly.”

Biases may be exacerbated by short discussion times, as reported by some respondents:

“Some reviewers spend too much time presenting their review, without focusing on the most important points, leaving less opportunity for discussion.”

“Discussions are usually too short, but tend to be OK.”

#### **Chair facilitation**

In terms of the usefulness of the chair in facilitating the application discussions, 68% of all reviewers reported that the chair’s involvement was either extremely useful or very useful (median of 2.0, 95%CI 1.9 to 2.1 on a scale of 1 to 5; Q4). Multiple respondents remarked specifically on how their chair helped or hindered the facilitation of the discussion and nearly all who commented on the chair recognized the importance of the chair’s role in discussion and scoring:

“The chairs of panels I have been on have been very important in directing, limiting, and policing the discussion. Most have done a poor job, even to the point of not cutting off inappropriate questions.”

“A good chair is absolutely essential to promoting balanced discussion, focusing debate, not letting debate draw on when it is clear differences of opinion are not going to be resolved based on discussion. If the chair is not good at this, the study section experience can be a miserable one.”

“An open-minded chair who is willing to direct discussion to key points is essential”

Several commented on the importance of the chair in facilitating participation:

“The chair of the panel is vital for success in the review process. Having a chair that encourages discussion and differing opinions is very important to having reviewers feel their voice has been heard, their opinion is valuable and promotes continuing grant review participation”

The overwhelming majority (88%, 95%CI 85% to 90%) of reviewers felt the discussions were fair and balanced, and mentioned the chair’s role in ensuring balanced discussion in their comments:

“Chair was great, fair and very well informed. It helped keep the discussion focused and helped adjust for extremes - overly laudatory reviews and

extremely negative reviews. Reviewers are variable. There were a couple of reviewers who were so negative I know how the applicant must feel when reading the review. Most were balanced.”

“I do know there is much effort made to provide fair and balanced discussion, though. It's most uncomfortable when one or two members of a panel, individuals who are more vociferous or opinionated, sway quieter reviewers who actually presented more logical reasons to support their scores. In other words, scores are sometimes based on the assertiveness of the reviewers' opinions rather than logic and rationale regarding the merits of the science. Those are situations where the Chair becomes extremely important but where I've seen applicants lose out.”

### Outcome influence

A total of 71% of reviewers agreed that panel discussion was extremely effective or very effective in influencing the outcome of the grant (median of 2.0, 95%CI 1.9 to 2.1 on a scale of 1 to 5; Q5). Nevertheless, many respondents suggested that there were relatively small scoring changes based on discussion:

“The intriguing thing for me is that after a very comprehensive and high-quality discussion, in many cases the preliminary scores do not change much”

and that if outcome is affected, it is more often in a negative than positive direction:

“Discussions rarely bring a grant to a better score, more often points out weaknesses. While I used to see that discussions brought folks to a consensus, I more often see it now as a veto of one because the only ones that achieve a fundable score are those without no negatives brought out in the discussion. This is the consequence of lots of good grants and continuing low funding lines.”

Importantly, some suggest an individual reviewer can have an undue influence on the outcome, either through dominating the discussion or through a poorly constructed initial review:

“the discussion and influence of a reviewer varies greatly and can make or break a borderline grant. The stronger reviewer will prevail. Few are willing to get into intellectual argument, especially if they haven't been assigned to the grant”

“I have found that when the primary and secondary reviewers disagree upon initial review, discussion

rarely changes the outcome much, even when one of the reviewers admits that they were wrong in down-scoring the application. And then the rest of the panel members split the difference. So, one grant with an excellent score by one reviewer and a mediocre score by the other ends up with a score outside the fundable range. And if the average of the scores places it out of the range to be discussed, the panel usually just lets it go so as to not increase the length of the meeting. So, a grant that one reviewer rated favorably can get torpedoed by a bad reviewer, even if that reviewer was totally off-base on their reasons for the bad score. Unfortunately, I have seen this from both sides - reviewer and grant applicant.”

Similarly, while 60% of reviewers definitely or most definitely agreed that the grant application discussions promoted the best science (median of 2.0, 95%CI 1.9 to 2.1 on a scale of 1 to 5; Q6), several respondents mentioned a level of conservatism in the discussion:

“Truly innovative grants are going to have an inherent risk. Panels often go for 'safe' bets.”

“The dynamics of panels are always interesting. There does seem to be some level of group-think, mostly resulting in more conservative review outcomes in my experience, but it's a complex interaction so difficult to describe accurately”

Again, respondents mentioned the disproportionate impact an individual bias can have on scoring and that unassigned reviewers may contribute significantly to bias:

“While most reviewers are knowledgeable and unbiased, it takes just one panel member to cast doubt on a grant.”

“I think the discussion with panel members who have not read the grant is pretty much a farce and can lead to dragging down of grants in an unfair manner.”

“In one or two cases where others were assigned to areas that I had greater expertise in, I felt their lack of expertise led to opinionated and influential comments that swayed the panel. In this situation, when I had not read the (unassigned) grant, I was able only to comment on the correctness of the statements of the reviewer, but not offer an alternative view based on knowledge of the science under discussion. In addition, there is pressure on reviewers

**Table 2** Discussion facilitation versus discussion affecting outcome

Question	Q5. Outcome affected (1 or 2)	Q5. Outcome unaffected (3, 4 or 5)	Significance (affected vs unaffected)
Q1. Did the grant application discussions facilitate reviewer participation?	Y = 96% [94–98%]	Y = 84% [79–89%]	$\chi^2 [1] = 30.8, p < 0.001, d = 0.33$
Q2. How useful were the grant application discussions in clarifying differing reviewer opinions?	2.0 [1.9–2.1]	3.0 [2.8–3.2]	$U[450, 184] = 67,562, p < 0.001, d = 0.86$
Q3. Was the format and duration of the grant application discussions sufficient to allow the non-assigned reviewers to cast well informed merit scores?	Y = 84% [76–88%]	Y = 65% [74–82%]	$\chi^2 [1] = 26.2, p < 0.001, d = 0.40$
Q4. How useful was the Chair in facilitating the application discussions?	2.0 [1.9–2.1]	3.0 [2.8–3.2]	$U[446, 181] = 54,003, p < 0.001, d = 0.78$

to make the evaluations short, which carries over to panel members not to belabor the discussion. I feel the time crunch has a negative impact on the fairness and thoroughness of the review.”

#### Differences based on panelist perception of outcome

Given the above responses, we were interested to explore whether views on panel discussion and outcome (Q5) were related to those of discussion effectiveness and quality (Q1–4). We separated respondents into 2 groups, those who felt the discussions affected the outcome (scoring 1 or 2 on Q5;  $N = 450$ ) and those who did not feel the discussions affected the outcome (scoring 3, 4, or 5 on Q5;  $N = 184$ ). We then compared the two groups in terms of responses surrounding discussion effectiveness and quality (Table 2). Significant differences were found between the two groups for all the questions, including views on reviewer participation (Q1), clarification of differing opinions (Q2), informing unassigned reviewers (Q3), and chair facilitation (Q4). In all cases, respondents who felt the outcome was affected by panel discussions viewed the discussion effectiveness and quality more favorably than those who felt the outcome was not affected by the discussions.

Respondent perceptions of discussion effectiveness and quality in terms of those who felt the discussions affected the outcome of the proposals (scored a 1 or 2 to this question) compared to those felt discussion was relatively unaffected (scored a 3, 4, or 5 to this question). Median values and 95% confidence intervals are displayed on the left and on the right are results from Mann-Whitney tests ( $U [n1, n2] = \text{value}, p = \text{value}$ ) or chi-square tests ( $\chi^2[\text{degree of freedom}] = \text{value}, p = \text{value}$ ). The calculated effect size ( $d$ ) is also provided.

#### Discussion

Our results indicate that, in general, reviewers felt that panel discussions were well facilitated across multiple dimensions, including favorable perceptions of panel inclusivity, leadership, and quality of communication. However, our results also indicated that reviewers who

did not think the discussion affects the outcome were much more likely to feel that several aspects of panel communication were problematic. Respondents from this survey mentioned uneven consideration of reviewers' opinions, low levels of participation from unassigned reviewers, and short discussion times as potential problems with panel discussions. These combined results support the idea that, at least in some instances, issues with the effectiveness and quality of panel discussion likely limit the influence of the discussion on panel scoring.

Similarly, while the majority of respondents felt the discussions affected the outcomes, several respondents commented how such discussions contained biases that limit the fairness of the review and its ability to select the best science. In particular, some reported how an individual reviewer (if not reined in) can have a greater than intended influence on the outcome, leading to a potential source of bias during discussion. Moreover, the manner of assignments (where only a few reviewers read the application) allows for this structure to bias the outcome, particularly for unassigned reviewers' scores. Several respondents also mentioned a level of conservatism in review panels with regards to innovative applications. While these suggestions of bias run counter to the majority of respondents who report that the discussion was fair and selected the best science, the presence of bias is still clearly an area of concern for some respondents. It may be that some reviewers are overconfident about the fairness of the discussion (potentially because they were directly involved in the discussion) and therefore not attuned to such biases [19]. Future studies could gather perceptions from impartial panel observers, such as scientific review officers who manage panels for funding agencies, to determine whether such perceptions of fairness are warranted.

Importantly, most respondents recognized the importance of the chair in facilitating the panel, and many comments suggested good chairs could help elicit reviewer participation, guide balanced discussion, and play important roles in modulating the length of discussion.

In addition, respondents noted the role of the chair in mitigating bias, limiting extreme reviewers, optimally leveraging panel expertise, encouraging the clear presentation of the assigned reviewer evaluations, and “directing discussion to key points.” Thus, based on the comments, there was almost universal support for the importance of the chair role in the facilitation of the panel to improve the impact of the discussion on the outcome while avoiding potential biases.

Nevertheless, nearly a third of respondents did not find the chair of their most recent panel to be very effective in facilitating the application discussions. If the chair is central to the effectiveness of panel discussion, more research should focus on identifying specific facilitative behaviors of effective chairs, and specific skills that moderate discussion in inclusive and unbiased ways. Future studies of discussion quality could assess for assertive and passive personality traits and panel leadership styles [14]. For instance, variability in discussion time may be a function of chair behavior (limit-setting versus allowing discussion). Further, the effectiveness of discussion from less persuasive reviewers may be hindered by a passive chair compared to a more engaged and assertive chair. Previous research has reported the importance of score-calibration comments and even laughter in the effectiveness of panel discussion, although it is unclear if these are affected by chair facilitation [8, 20, 21]. Future studies should include a focus on the social influences and group dynamics between panel reviewers, informed by the literature on small group decision making in other contexts.

Our results also suggest that reviewers are unlikely to participate in discussions of proposals on which they are not the assigned reviewers, likely due to the fact that they do not read these proposals. According to some respondents, bias can result from unassigned reviewer reliance on only the discussion to inform them about proposals’ strengths and weaknesses. This model of panel scoring, where most panelists score the proposal without having read it, may not achieve the goal of leveraging panel expertise. Because many reviewers are overburdened [18], solutions to achieve this goal should be explored, such as examining the optimal number of assigned reviewers [22], or asking unassigned reviewers to read proposal abstracts and critiques ahead of the meeting.

It should be mentioned that one potential limitation to this study is the relatively low response rate (6.7%) for this sample, although this rate is similar to those in other recent surveys on journal peer review [23–25]. Furthermore, the demographics of our sample are very similar to those of NIH study section members, according to recent reports [26]. Additionally, comparing the larger, full sample of incomplete responses ( $n = 1231$ ) to

the one used in this manuscript, we find very similar demographics, which suggests that this sample is representative of the larger population. Another potential limitation of the study is that the subset of the sample who offered comments related to review outcomes (that the discussion influenced the outcome and promoted the best science) expressed negative sentiments and had more negative median ratings to these questions than those who did not make a comment. This pattern of data may reflect negative response bias among those who offered comments related to the questions about review outcomes, such that their comments may not be representative of the entire sample. However, the themes of these comments have been noted in other studies, namely, the lack of score shifting after discussion [9, 13] and conservatism on panels with regards to lack of support for innovative work [17, 27].

## Conclusions

Overall, our results find that most reviewers think the quality and effectiveness of panel discussion is high and does influence the outcome of the review. Conversely, our results also point to poor panel facilitation as a potential factor that limits the influence the discussion has on scoring and may even introduce biases. It is also clear from this study that reviewers feel a strong chair can help to avoid such biases and ensure engagement and inclusion; therefore, it is of great importance that future chairs should be properly trained in how to lead and facilitate a discussion. Moreover, future review processes could be informed by the science of leadership and team communication to enhance consistency, inclusivity, and impartiality in panel discussions [14, 28–31].

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s41073-020-00093-0>.

**Additional file 1.** Supporting Information - Peer Review Survey.

**Additional file 2.** Supplemental Table 1 - Comparison of Answers of Early Versus Late Respondent Groups.

## Acknowledgements

We thank all the survey respondents for taking the time to provide their thoughts and opinions on peer review.

## Authors’ contributions

Authors’ contributions: SAG, KBS, LAT, and SRG had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. SAG, KBS, and LAT contributed to study concept and design; KBS prepared the Institutional Review Board application and was its point of contact; SAG and LAT exported and prepared the data; all authors contributed to the analysis and interpretation of the data; SAG drafted the manuscript, KBS substantially revised it, LAT and SRG contributed to the critical revision of the manuscript; and SAG provided study supervision. All authors read and approved the final manuscript.



### Funding

The author(s) received no specific funding for this work.

### Availability of data and materials

A full copy of the peer review survey is available in the **S1 File in the Supporting Information**. The raw, anonymized data are also available on Figshare (<https://doi.org/10.6084/m9.figshare.8132453.v1>).

### Ethics approval and consent to participate

The study was reviewed by the Washington State University Office of Research Assurances (Assurance# FWA00002946) and granted an exemption from IRB review consistent with 45 CFR 46.101(b)(2). Participants were free to choose whether or not to participate in the survey and consented by their participation.

### Consent for publication

Not applicable.

### Competing interests

The authors declare they have no competing interests.

### Author details

<sup>1</sup>Scientific Peer Advisory and Review Services, American Institute of Biological Sciences, Herndon, VA, USA. <sup>2</sup>Washington State University, Vancouver, WA, USA.

Received: 11 December 2019 Accepted: 24 April 2020

Published online: 15 May 2020

### References

1. NIH. Peer Review. 2018. Available from: <https://grants.nih.gov/grants/peerreview22713webv2.pdf>. Accessed March 2020.
2. Liaw L, Freedman JE, Becker LB, Mehta NN, Liscum L. Peer review practices for evaluating biomedical research grants: A scientific statement from the American Heart Association. *Circ Res*. 2017;121(4):e9–e19.
3. Cicchetti DV. The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behav Brain Sci*. 1991;14(1): 119–35.
4. Cole S, Simon GA. Chance and consensus in peer review. *Science*. 1981; 214(4523):881–6.
5. Fogelholm M, Leppinen S, Auvinen A, Raitanen J, Nuutinen A, Väänänen K. Panel discussion does not improve reliability of peer review for medical research grant proposals. *J Clin Epidemiol*. 2012;65:47–52.
6. Pier EL, Brauer M, Filut A, Kaatz A, Raclaw J, Nathan MJ, et al. Low agreement among reviewers evaluating the same NIH grant applications. *Proc Natl Acad Sci*. 2018;115(12):2952–7.
7. Brown AL, Ash D, Rutherford M, Nakagawa K, Gordon A, Campione JC. Distributed expertise in the classroom. In: Salomon G, editor. *Distributed cognitions: Psychological and educational considerations*. Cambridge: Cambridge University Press; 1993. p. 188–228.
8. Pier EL, Raclaw J, Kaatz A, Brauer M, Carnes M, Nathan MJ, et al. ‘Your comments are meaner than your score’: Score calibration talk influences intra-and inter-panel variability during scientific grant peer review. *Res Eval*. 2017;26(1):1–14.
9. Carpenter AS, Sullivan JH, Deshmukh A, Glisson SR, Gallo SA. A retrospective analysis of the effect of discussion in teleconference and face-to-face scientific peer-review panels. *BMJ Open*. 2015;5(9):e009138.
10. Martin MR, Kopstein A, Janice JM. An analysis of preliminary and post-discussion priority scores for grant applications peer reviewed by the Center for Scientific Review at the NIH. *PLoS One*. 2010;5:e13526.
11. Obrecht M, Tibelius K, D’Aloisio G. Examining the value added by committee discussion in the review of applications for research awards. *Res Eval*. 2007;16:79–91.
12. Fleurence RL, Forsythe LP, Lauer M, Rotter J, Ioannidis JP, Beal A, et al. Engaging patients and stakeholders in research proposal review: The patient-centered outcomes research institute. *Ann Intern Med*. 2014;161: 122–30.
13. Pier EL, Raclaw J, Nathan MJ, Kaatz A, Carnes M, Ford CE. Studying the study section: How group decision making in person and via videoconferencing affects the grant peer review process. W CER Working Paper No. 2015-6. Wisconsin Center for Education Research; 2015. 24 p.
14. National Research Council. *Enhancing the effectiveness of team science*. Washington, DC: The National Academies Press, 2015. 280 p.
15. Langfeldt L. The decision-making constraints and processes of grant peer review, and their effects on the review outcome. *Soc Stud Sci*. 2001;31(6): 820–41.
16. NIH CSR. Reviewer quick feedback survey results. 2015. Available from: <https://public.csr.nih.gov/sites/default/files/2017-10/ReviewerQuickFeedbackSurveyResults.pdf>.
17. Gallo S, Thompson L, Schmalig K, Glisson S. Risk evaluation in peer review of grant applications. *Environ Syst Decis*. 2018;38:216–29.
18. Gallo SA, Thompson LA, Schmalig KB, Glisson SR. Participation and motivations of grant peer reviewers: A comprehensive survey. *Sci Eng Ethics*. 2019. <https://doi.org/10.1007/s11948-019-00123-1>.
19. Moore DA, Healy PJ. The trouble with overconfidence. *Psychol Rev*. 2008; 115(2):502–17.
20. Raclaw J, Ford CE. Laughter and the management of divergent positions in peer review interactions. *J Pragmat*. 2017;113:1–15.
21. Pier EL, Raclaw J, Carnes M, Ford CE, Kaatz A. Laughter and the chair: Social pressures influencing scoring during grant peer review meetings. *J Gen Intern Med*. 2019;34(4):513–4.
22. Snell RR. Menage a quoi? Optimal number of peer reviewers. *PLoS One*. 2015;10(4):e0120838.
23. Sense About Science. Peer review survey 2009. Available from: <http://archive.senseaboutscience.org/pages/peer-review-survey-2009.html>. Accessed March 2020.
24. Ware M. Peer review: Benefits, perceptions and alternatives. *Publishing Research Consortium* 2008. Available from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.214.9676&rep=rep1&type=pdf> Accessed March 2020.
25. Ware M, Monkman M. *Peer review in scholarly journals: Perspective of the scholarly community—An international study*. London, UK: Publishing Research Consortium; 2008. Available from: <https://pdfs.semanticscholar.org/7050/9ee2d48f00b0b5ed1d471773f6c0d9b45b30.pdf>.
26. NIH OER. *Enhancing peer review survey results report*. 2013. Available from: [https://enhancing-peer-review.nih.gov/docs/Enhancing\\_Peer\\_Review\\_Report\\_2012.pdf](https://enhancing-peer-review.nih.gov/docs/Enhancing_Peer_Review_Report_2012.pdf).
27. Boudreau KJ, Guinan EC, Lakhani KR, Riedl C. Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. *Manage Sci*. 2016;62(10):2765–83.
28. Bos N, Olson J, Gergle D, Olson G, Wright Z. Effects of four computer-mediated communications channels on trust development. *CHI*. 2002:135–40.
29. Chen G, Farh JL, Campbell-Bush EM, Wu Z, Wu X. Teams as innovative systems: Multilevel motivational antecedents of innovation in R&D teams. *J Appl Psychol*. 2013;98(6):1018–27.
30. Driskell JE, Radtke PH, Salas E. Virtual teams: Effects of technological mediation on team performance. *Group Dyn*. 2003;7:297–323.
31. Kozlowski SW, Ilgen DR. Enhancing the effectiveness of work groups and teams. *Psychol Sci Public Interest*. 2006;7(3):77–124.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

