Comparative Data on Face-to-face, Virtual, and Hybrid Format Peer Review Meetings– Reviewer Surveys and Quantitative Meeting Measures

Office of the Director

Executive Summary

This report examines the impact of hybrid meeting format on peer review processes. It compares hybrid meetings, those that have both in-person and remote attendees, to face-to-face meetings and to virtual meetings. The Center for Scientific Review (CSR) manages the peer review of over 60,000 NIH grant applications annually. When the COVID-19 pandemic began in March 2020 CSR shifted its review meetings online, using the Zoom.gov video meeting platform. Face-to-face meetings were resumed in 2022, with panels holding one of every three annual meetings in-person and the remainder online. Beginning in 2023, CSR implemented hybrid meetings (combined face-to-face and virtual) in addition to the one of three in-person schedule. Thus, from 2022 forward review meetings held during the same period were run using different meeting formats. This created a natural experiment for evaluating the effects of meeting format on peer review processes and outcomes. It is one of a series of reports stemming from the Center for Scientific Review's ongoing effort to understand the impact of meeting format on the peer review process and focuses on hybrid meetings. Prior reports examined virtual versus face=to-face meetings (2022 data, 2023 data).

CSR collected two types of data to evaluate outcomes: participant survey data and multiple quantitative meeting measures. Survey data were designed to assess reviewers' observations of the quality of the review and meeting experiences. Quantitative meeting measures were selected to evaluate roster characteristics and scoring practices. Uniform data were collected for the reports on face-to-face and virtual meetings referenced above, permitting examination of the effects of the hybrid format on outcomes. To facilitate comparison with other CSR reports on the effects of meeting format, data from reviewers at prior face-to-face and virtual meetings from the Summer 2023 review meetings are included in some figures for comparison.

Main Findings:
The peer review process functions well in all three meeting formats (virtual, face-to-face, hybrid). Reviewers give high ratings to the panel's ability to prioritize the highest impact proposals, and to the overall quality of review, regardless of format. There are differences, generally small. Broadly, in-person meetings are rated better than hybrid, and hybrid are rated better than virtual. For example, we see this pattern on ratings of reviewer engagement and quality of discussion. Meeting in a hybrid format appears to improve the meeting experience for those attending online, while giving in-person attendees an experience comparable to attendees at in-person meetings. In examining scoring data and characteristics of who is recruited to attend meetings in person or remotely, meeting format appears to have little effect. Finally, most reviewers prefer meeting in person, although a minority prefer virtual or hybrid meetings.

Introduction

This report examines the impact of hybrid meeting format on peer review processes. It compares hybrid meetings, those that have both in-person and remote attendees, to face-to-face meetings and to virtual meetings. It is one of a series of reports stemming from the Center for Scientific Review's ongoing effort to understand the impact of meeting format on the peer review process. The Center for Scientific Review (CSR) manages the peer review of over 60,000 NIH grant applications annually. When the COVID-19 pandemic began in March 2020, CSR shifted its review meetings online, using the Zoom.gov video meeting platform. During Fall 2022, CSR reimplemented face-to-face meetings such that each standing study section and recurring special emphasis panel (handling fellowship or small business applications) shifted to a schedule of one face-to-face and two virtual meetings each year. Thus, in each round, many meetings would be held virtually while many others were held face-to-face. The comparison and analysis of [Fall 2022 meetings](#) (January 2023 Council Round) by meeting format was followed by an [additional report](#) for the Summer 2023 meetings (August and October Council round). During Summer and Fall 2023, and Winter and Summer 2024, CSR held a small number of review meetings in a hybrid format. An interim report focused on reviewer surveys was [posted](#) earlier this year – the present report adds survey data on 27 hybrid meetings from the Summer 2024 review meetings, as well as quantitative meeting measures (e.g. data on scoring).

CSR collected two types of data to evaluate outcomes: participant survey data and multiple quantitative meeting measures. Survey data were designed to assess reviewers' observations of the quality of the review and meeting experiences. Quantitative meeting measures were selected to evaluate roster characteristics and scoring practices. Uniform data were collected for the reports on face-to-face and virtual meetings referenced above, permitting examination of the effects of the hybrid format on outcomes. Thus, to facilitate comparison with other CSR reports on the effects of meeting format, data from reviewers at prior face-to-face and virtual meetings from the Summer 2023 review meetings are included in some figures for comparison.

A note on terminology: *Meeting formats* are referred to as face-to-face, virtual, or hybrid, whereas *how a reviewer participated* in a hybrid meeting (type of attendee) is designated as either in-person or remote. All attendees at virtual meetings attend remotely; virtually all attendees at face-to-face meetings attend in person. Please see [Appendix A](#) for more information.

Main Findings:

- **Overall quality of review.** Reviewers rate the quality of CSR review meetings highly, regardless of meeting format. In virtual meetings, 94% of reviewers rated the overall quality of their review meeting as excellent or good, and ratings were even higher for other formats. There was no difference in ratings of overall quality between face-to-face and hybrid review meetings.
- **Quality of discussions and reviewer engagement.** Discussions and reviewer engagement were rated better for face-to-face meetings than hybrid meetings, and hybrid were rated better than virtual.
- **Panel ability to prioritize applications.** Ratings for face-to-face and hybrid meetings did not differ, and both were better than ratings of virtual meetings.
- **Reviewer experience and participation**. Face-to-face and hybrid meetings consistently rate better than virtual meetings.
- **Attention span**. Reviewers in face-to-face and hybrid meetings rated their attention better than did the reviewers in virtual meetings. Face-to-face meeting attendees rated their attention slightly better than hybrid attendees.
- **Reviewers' meeting format preferences**. Overall, reviewers prefer face-to-face meetings to virtual meetings (60 to 22%) with the remainder having no preference.

- **Effect of meeting format on rosters**. Virtual or hybrid meeting formats do not change who attends – they do not draw more underrepresented minorities (URMs), women, early career reviewers, or reviewers from places distant from the Washington, D.C. area.
- **Effects of meeting format on scoring**. There are statistically significant, very small effect size, differences between meeting formats on some scoring parameters.
- **Focus on mode of attendance.** Ratings often differed between those who attended hybrid meetings in person versus remotely.
    - In-person attendees rated the overall quality of the review higher and gave better ratings on measures of discussion quality and their experience as reviewers than did remote attendees.
    - Compared to virtual attendees, remote reviewers gave better ratings to the productivity of discussion, reviewer engagement, panel ability to prioritize applications, and overall quality of review.
    - Compared to virtual attendees, remote reviewers rated their attention span better.
    - Face-to-face meeting attendees and in-person attendees at hybrid meetings did not differ on ratings of productivity of discussion and overall quality of review; the face-to-face meeting attendees did rate reviewer engagement slightly better.
    - Compared to face-to-face meeting attendees, in-person attendees at hybrid meetings said they contributed to discussion more and felt more confident voicing their opinions.

In the first part of the report that follows, we present the survey data, first comparing attendee experiences between meeting formats (face-to-face, virtual, and hybrid), followed by a closer examination of attendee experiences in hybrid meetings – in whether they attended the hybrid meeting in person or remotely.

The second part of the report details quantitative meeting measures, first by meeting characteristics (application counts, roster sizes, reviewer workload trends) and then by roster composition – this is reported at the level of meeting formats (face-to-face, virtual, and hybrid). The last component of the report focuses on scoring behaviors, first between meeting formats, followed by a closer examination of scoring behaviors in hybrid meetings only by whether reviewers attended the hybrid meetings in person or remotely.

Overall Meeting Format – Survey Data and Analyses
See *Appendix B* for detailed methods.

Results

The survey was administered to 8,443 reviewers who attended review meetings in the Summer and Fall of 2023 and the Winter and Summer of 2024. The survey was completed by 4,042 reviewers, for a response rate of 48%. Of the respondents, 45% attended virtual meetings (n = 1,707), 30% attended face-to-face meetings (n = 1,167), and 25% attended hybrid meetings (n = 959)[1,2]. Among those who attended hybrid meetings, 47% (n = 455) attended remotely and 53% (n = 504) attended in person. See Table 1 for reviewer characteristics.

| Table 1. Reviewer Characteristics of Survey Respondents	
Reviewer Characteristics	**% Survey Respondents (n = 3,833)**
Gender	
Male	52
Female	45
Withheld	3
Race	
American Indian or Alaska Native	< 1
Asian	22
Black or African American	4
More than one race	2
Native Hawaiian or Pacific Islander	< 1
White	63
Withheld	9
Ethnicity	
Hispanic/Latino	9
Non-Hispanic	86
Withheld	5
URM	
No	78
Yes	14
Withheld	8
Career Stage	
Professor	45
Associate Professor	32
Assistant Professor	18

---

[1] To increase sample size, participants that attended hybrid meetings came from four different rounds of review meetings—155 participants attended the Summer 2023 meetings, 95 participants attended the Fall 2023 meetings, 287 participants attended the Winter 2024 meetings, and 437 participants attended the Summer 2024 meetings, for a total of 974 hybrid responses.
[2] For reviewers who responded to more than one survey, the first survey response(s) were excluded to avoid non-independent data while retaining as much hybrid data as possible--resulting in the removal of 209 cases/survey responses (primarily responses from reviewers in face-to-face and virtual meetings from Summer 2023) and an adjusted sample size of 3,833.

	Other	5

Overall Meeting Format Preferences

- Figure 1 shows the meeting format preference of reviewers broken down by how they participated in the meeting[3].

- Reviewers overall prefer face-to-face meetings, but the degree of preference varies based on reviewers' meeting format.

- Reviewers in face-to-face meetings were more likely to prefer face-to-face meetings than reviewers in virtual and hybrid meetings; reviewers in virtual meetings were more likely to prefer virtual meetings than reviewers in face-to-face meetings; reviewers in hybrid meetings were more likely to prefer face-to-face meetings than reviewers in virtual meetings and were more likely to prefer virtual meetings than reviewers in face-to-face meetings.
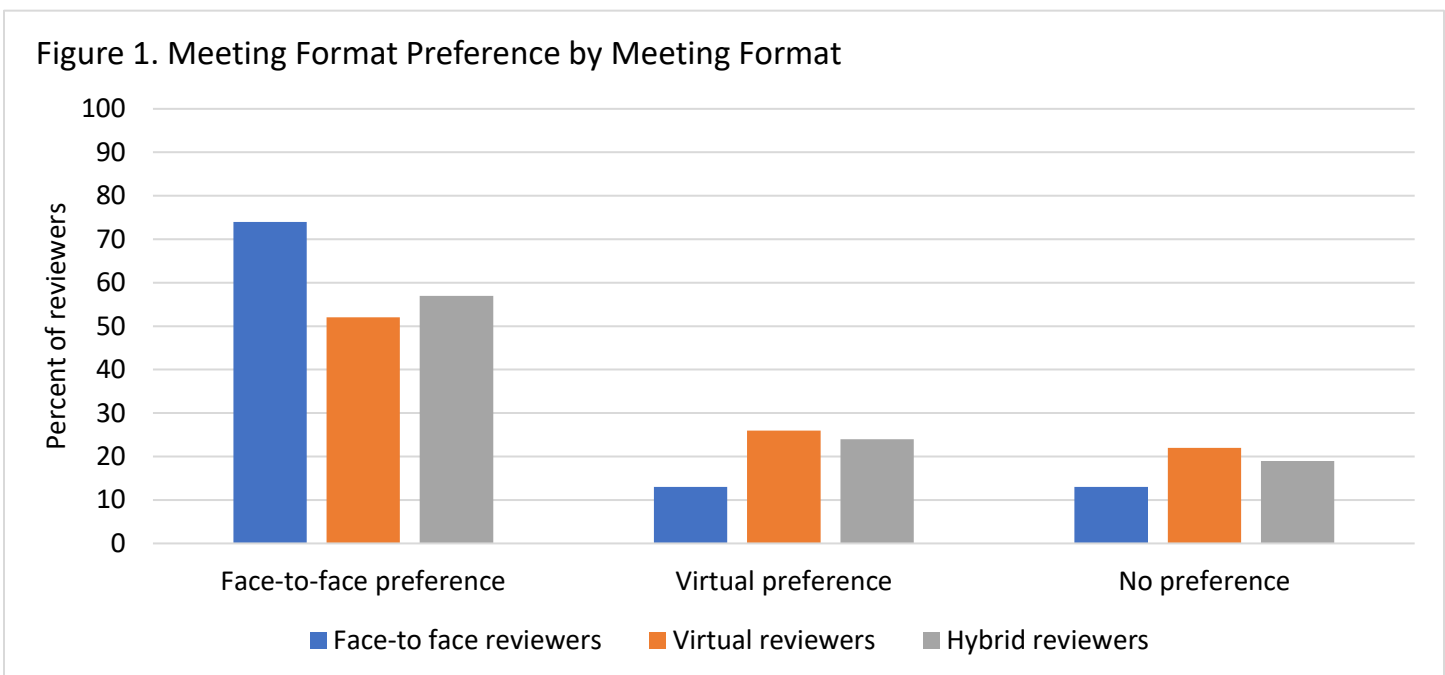
Figure 1. Meeting Format Preference by Meeting Format

There was a significant association between the format of reviewers' meetings and their meeting format preferences χ^2 (4) = 143.093, $p < .001$, with a small effect size (ϕ_c = .193, $p < .001$). Post hoc z-tests show that 1) the proportion of reviewers in face-to-face meetings that preferred face-to-face meetings was significantly more than the proportion of reviewers in virtual and hybrid meetings that preferred face-to-face meetings, 2) the proportion of reviewers in hybrid meetings that preferred face-to-face meetings was significantly more than the proportion of reviewers in virtual meetings that preferred face-to-face meetings, 3) the proportion of reviewers in face-to-face meetings that preferred virtual meetings or had no meeting preference was significantly less than the proportion of reviewers in virtual and hybrid meetings that preferred virtual meetings or had no meeting preference, 4) there was no significant differences between the proportion of reviewers in virtual and hybrid meetings that preferred virtual meetings or had no meeting preference.

---

[3] Note that meeting formats were not assigned randomly to review meetings or reviewers. CSR's post-pandemic plan, the preferences of scientific review officers, and reviewers' preferences were all taken into consideration when determining the meeting format and the attendee type for each reviewer. Thus, these findings could reflect those reviewers who, a priori, preferred to attend face-to-face or virtually.

Quality of Review for all Meetings

- Figure 2 shows reviewer perceptions of the quality of the review meeting by meeting format.

- Over 90% of all reviewers thought the overall quality of the meetings was good or excellent. There was no significant difference in ratings of overall quality between face-to-face and hybrid meetings; both were rated better than virtual meetings. (Effect size was small; see Table 2)

- Discussions and reviewer engagement were rated better for face-to-face meetings than hybrid meetings, and hybrid were rated better than virtual (see Table 2). The effect size of the face-to-face and hybrid versus virtual meetings comparison for discussion quality was medium and was the greatest difference observed in any comparison.
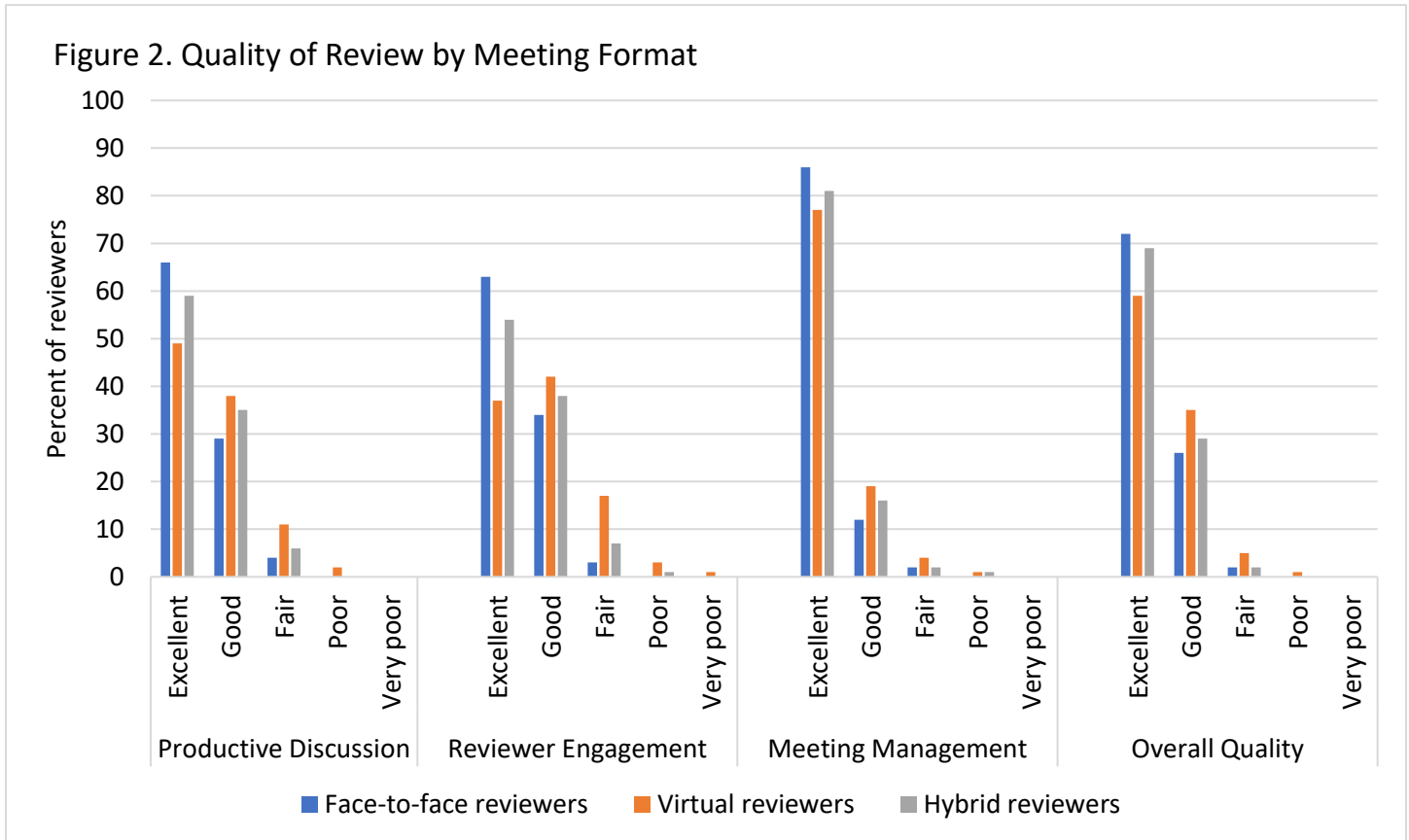
Figure 2. Quality of Review by Meeting Format

Table 2. Reviewers' Assessment of the Quality of the Review by Meeting Format			
	Face-to-face Meetings (M, SE)	Virtual Meetings (M, SE)	Hybrid Meetings (M, SE)	ANOVA statistic and *est.* ω^2
Productive Discussions	4.61 (.02)	4.34 (.02)	4.53 (.02)	Welch's $F(2, 2374.858) = 57.525$, $p < .001$; *est.* $\omega^2 = .03$
A Games-Howell post hoc test revealed that the review discussion was significantly better for face-to-face ($p < .001$, $d = .40$) and hybrid ($p < .001$, $d = .27$) meeting formats compared to virtual meeting formats. There was also a significant difference in the quality of the discussion between face-to-face and hybrid meeting formats ($p = .006$, $d = .13$).				
Reviewer Engagement	4.58 (.02)	4.13 (.02)	4.45 (.02)	Welch's $F(2, 2382.049) = 149.597$, $p < .001$; *est.* $\omega^2 = .07$
A Games-Howell post hoc test revealed that reviewer engagement was significantly better for face-to-face ($p < .001$, $d = .62$) and hybrid ($p < .001$, $d = .43$) meeting formats compared to virtual meeting formats. There was also a significant difference in reviewer engagement between face-to-face and hybrid meeting formats ($p < .001$, $d = .21$).				
Meeting Management	4.84 (.01)	4.70 (.01)	4.79 (.02)	Welch's $F(2, 2358.427) = 24.340$, $p < .001$; *est.* $\omega^2 = .01$
A Games-Howell post hoc test revealed that the meeting management was significantly better for face-to-face (p < .001, d = .27) and hybrid (p < .001, d = .16) meeting formats compared to virtual meeting formats. There was also a significant difference in meeting management between face-to-face and hybrid meeting formats ($p = .044$, $d = .11$).				
Overall Quality of Review	4.70 (.02)	4.53 (.02)	4.66 (.02)	Welch's $F(2, 2350.916) = 34.382$, $p < .001$; *est.* $\omega^2 = .02$
A Games-Howell post hoc test revealed that the overall quality of the review was significantly better for face-to-face ($p < .001$, $d = .30$) and hybrid ($p < .001$, $d = .22$) meeting formats compared to virtual meeting formats. There was no significant difference in the overall quality of the review between face-to-face and hybrid meeting formats ($p = .199$).				

Personal Meeting Experience and Participation

- Figure 3 shows reviewer ratings of their own experience and participation at the review. Overall, around 90% of all reviewers stated that they always or often felt confident voicing their opinions, felt others were responsive and receptive to their feedback, and were able to clearly communicate their opinions.

- Reviewers in face-to-face and hybrid meetings rated their meeting experience significantly better than did reviewers in virtual meetings: they said they contributed to discussion more, felt more confident voicing opinions, and found others more responsive to feedback when compared to reviewers in virtual meetings. Effect sizes were small. See Table 3.

- Except for opinions on comfort voting outside the range, reviewers who attended face-to-face and hybrid meetings reported a significantly better meeting experience and a significantly higher level of participation than those who attended virtual meetings—although the magnitude of these effects is small (see Table 3).

- **There were no significant differences in reviewer ratings between face-to-face and hybrid meetings on any of these measures**.
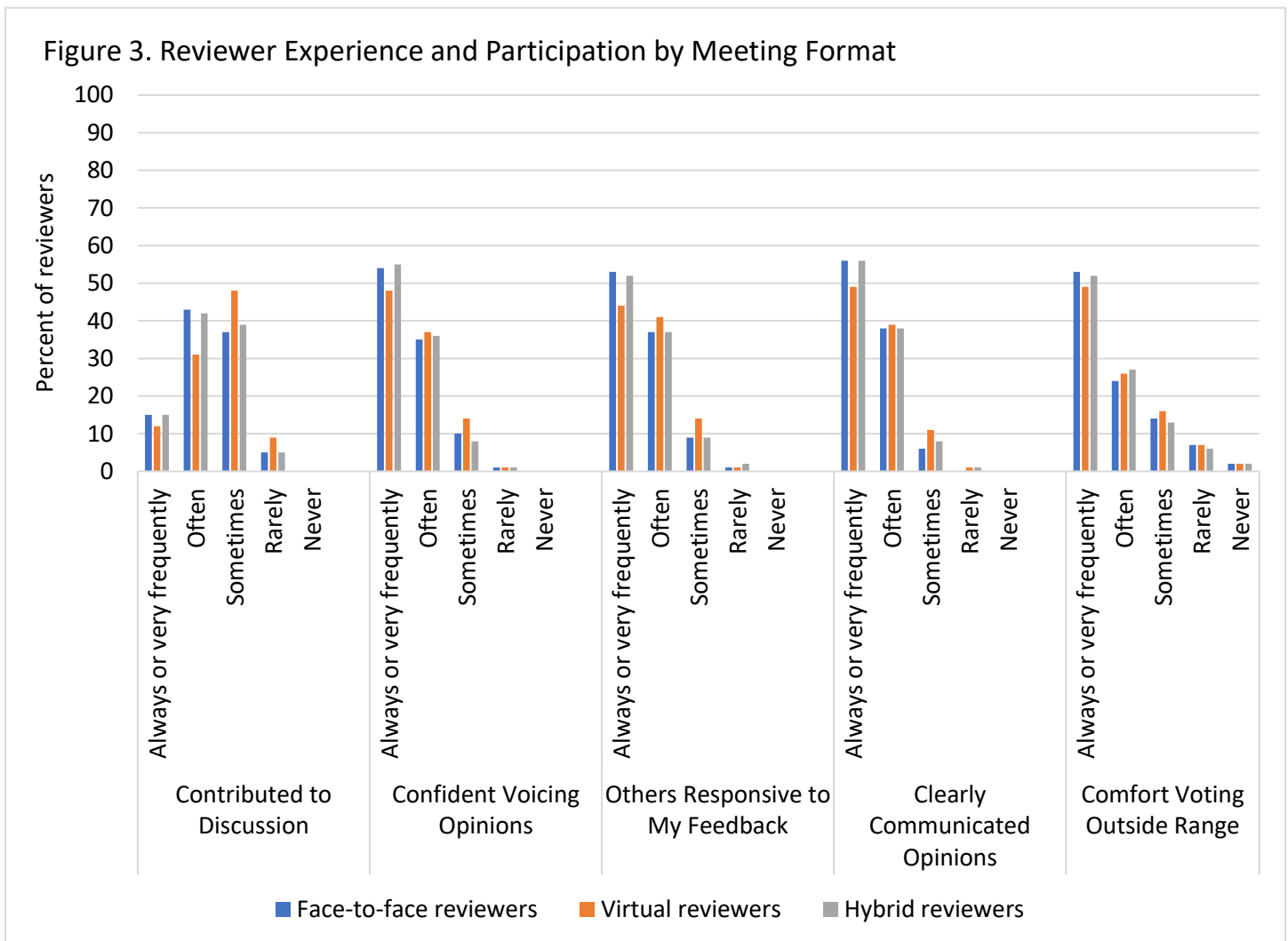

Figure 3. Reviewer Experience and Participation by Meeting Format

| Table 3. Reviewers' Meeting Experience and Participation by Meeting Format				
	Face-to-face Meetings (*M, SE*)	Virtual Meetings (*M, SE*)	Hybrid Meetings (*M, SE*)	ANOVA statistic and *est. ω^2*
Contributed to Discussion	3.69 (.02)	3.46 (.02)	3.66 (.03)	Welch's $F(2, 2284.295) = 35.133$, $p < .001$; $\omega^2 = .02$
A Games-Howell post hoc test revealed that reviewers' contributions to the discussion were significantly more at face-to-face ($p < .001$, $d = .28$) and hybrid ($p < .001$, $d = .25$) meeting formats compared to virtual meeting formats. There was no significant difference in the quality of the discussion between face-to-face and hybrid meeting formats ($p = .739$).				
Confident Voicing Opinions	4.41 (.02)	4.31 (.02)	4.44 (.02)	Welch's $F(2, 2285.472) = 11.520$, $p < .001$; *est.* $\omega^2 = .00$
A Games-Howell post hoc test revealed that reviewers felt significantly more confident voicing their opinions at face-to-face ($p = .003$, $d = .13$) and hybrid ($p < .001$, $d = .18$) meeting formats compared to virtual meeting formats. There was no significant difference in reviewers' confidence voicing their opinions between face-to-face and hybrid meeting formats ($p = .432$).				
Others Receptive and Responsive to Feedback	4.43 (.02)	4.27 (.02)	4.39 (.02)	$F(2, 3781) = 17.514$, $p < .001$; $\omega^2 = .00$
A Games-Howell post hoc test revealed that reviewers who attended face-to-face ($p < .001$, $d = .22$) and hybrid ($p < .001$, $d = .16$) meeting formats thought others were significantly more receptive and responsive to their feedback compared to reviewers who attended virtual meeting formats. There was no significant difference in reviewers' perceptions of others being receptive and responsive to their feedback between face-to-face and hybrid meeting formats ($p = .541$).				
Clearly Communicated Opinions	4.50 (.02)	4.36 (.02)	4.47 (.02)	Welch's $F(2, 2285.911) = 15.588$, $p < .001$; *est.* $\omega^2 = .00$
A Games-Howell post hoc test revealed that reviewers believed that they were significantly more able to clearly communicate their opinions at face-to-face ($p < .001$, $d = .21$) and hybrid ($p < .001$, $d = .16$) meeting formats compared to virtual meeting formats. There was no significant difference in reviewers' perceptions of their ability to clearly communicate their opinions between face-to-face and hybrid meeting formats ($p = .581$).				
Comfortable Voting Outside Range	4.20(.03)	4.13 (.03)	4.22 (.04)	$F(2,3254) = 2.375$, $p = .093$
A Games-Howell post hoc test revealed no significant differences in reviewers comfort voting outside the range between face-to-face ($p = .217$) and hybrid ($p = .124$) meeting formats compared to virtual meeting formats or between hybrid and face-to-face meeting formats ($p = .944$).				

Reviewers Attention Span for all Meetings

- Figure 4 shows reviewers' report of their attention throughout the meeting, by meeting format.
- Overall, reviewers who attended face-to-face and hybrid meetings had significantly longer attention spans than did reviewers who attended virtual meetings. The magnitude of these effects was small to medium. Reviewers in face-to-face meetings rated their attention significantly better than reviewers in hybrid meetings—the magnitude of this effect is small.

Figure 4. Reviewers Attention Span by Meeting Format
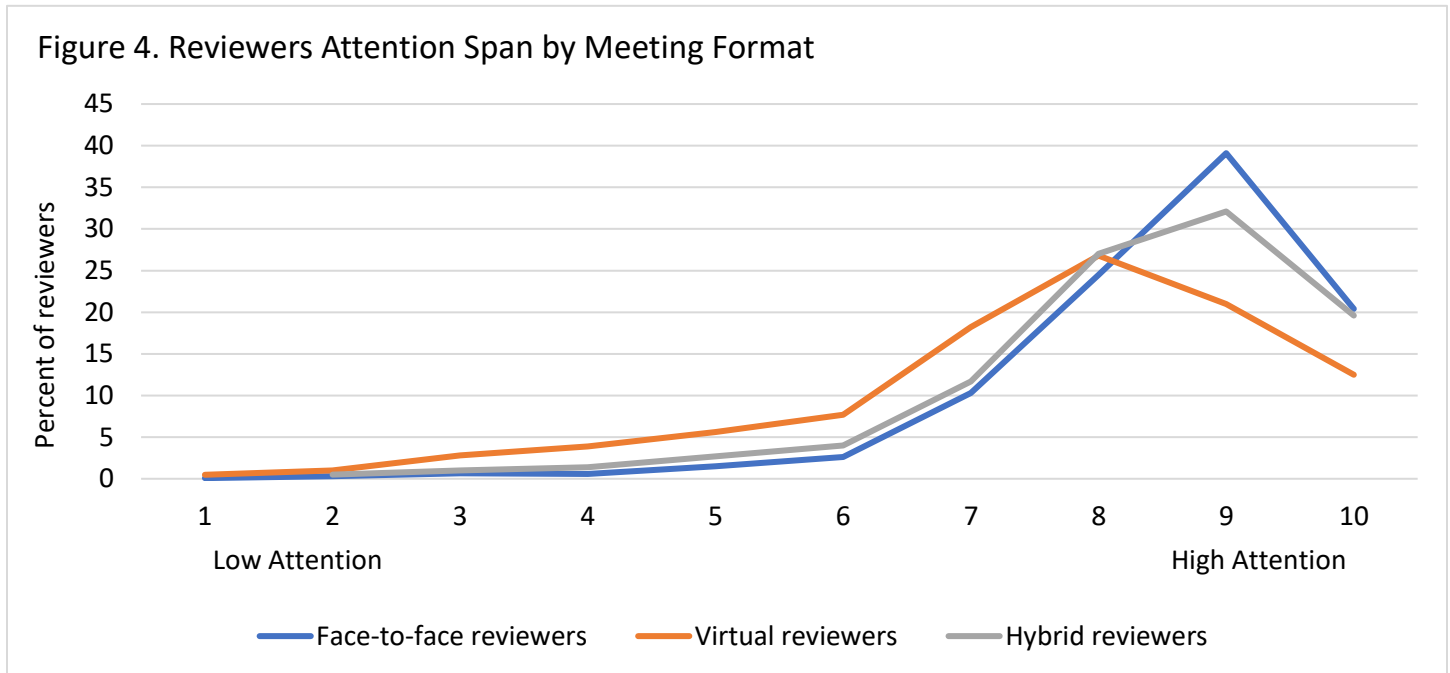
There was a significant effect of meeting format on reviewers attention span, Welch's $F(2, 2374.069) = 135.251$, $p < .001$, $est.\ \omega^2 = .07$. A Games-Howell post hoc test revealed that reviewers' attention at the meeting was 1) significantly more for face-to-face ($p < .001$, $d = .60$) and hybrid ($p < .001$, $d = .44$) meetings compared to virtual meeting formats and 2) significantly more for face-to-face meetings compared to hybrid meeting formats ($p = .001$, $d = .16$).

Prioritizing Applications

- Figure 5 shows data capturing reviewers' perceptions of the panel's ability to prioritize applications. Over 90% of all reviewers believed (said "strongly agree" or "agree") that the panel was able to prioritize applications according to their impact and scientific merit.
- Reviewers who attended face-to-face and hybrid meetings rated the panel's ability to prioritize applications significantly better than did those who attended virtual meetings—the magnitude of this effect is small.
- **Ratings of reviewers' ability to prioritize applications did not differ between hybrid and face-to-face meetings**.

Figure 5. Reviewers' Ability to Prioritize Applications by Meeting Format
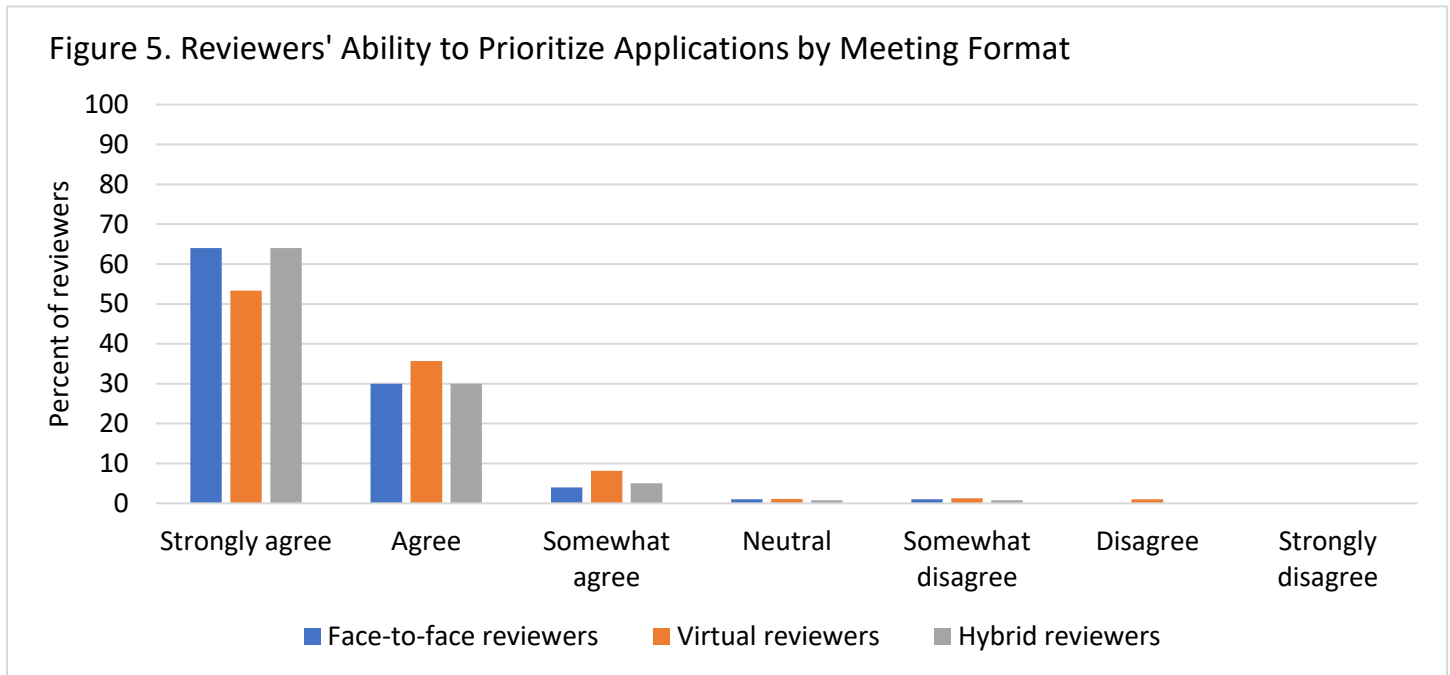
There was a significant effect of meeting format on reviewers ability to prioritize applications according to their impact and scientific merit, Welch's $F(2, 2353.455) = 24.365$, $p < .001$, $est.\ \omega^2 = .01$. A Games-Howell post hoc test revealed that reviewers were better able to prioritize applications at face-to-face ($p < .001$, $d = .22$) and hybrid ($p < .001$, $d = .24$) meetings compared to virtual meetings. There were no significant differences in the reviewer's ability to prioritize applications between face-to-face and hybrid meeting formats ($p = .916$).

Did Discussions Help Evaluations?

- Figure 6 shows data capturing reviewers' perceptions of the panel's ability to discuss and evaluate applications. Over 90% of all reviewers believed (said "strongly agree" or "agree") that the scientific discussions helped the panel evaluate the applications being reviewed.
- Reviewers who attended face-to-face and hybrid meetings believed that the discussions were significantly more helpful for evaluating the applications than did those who attended virtual meetings—the magnitude of this effect is small.
- **There was no significant difference in ratings on this item between reviewers who attended face-to-face versus hybrid meetings.**

Figure 6. Discussions Helped with Evaluations by Meeting Format
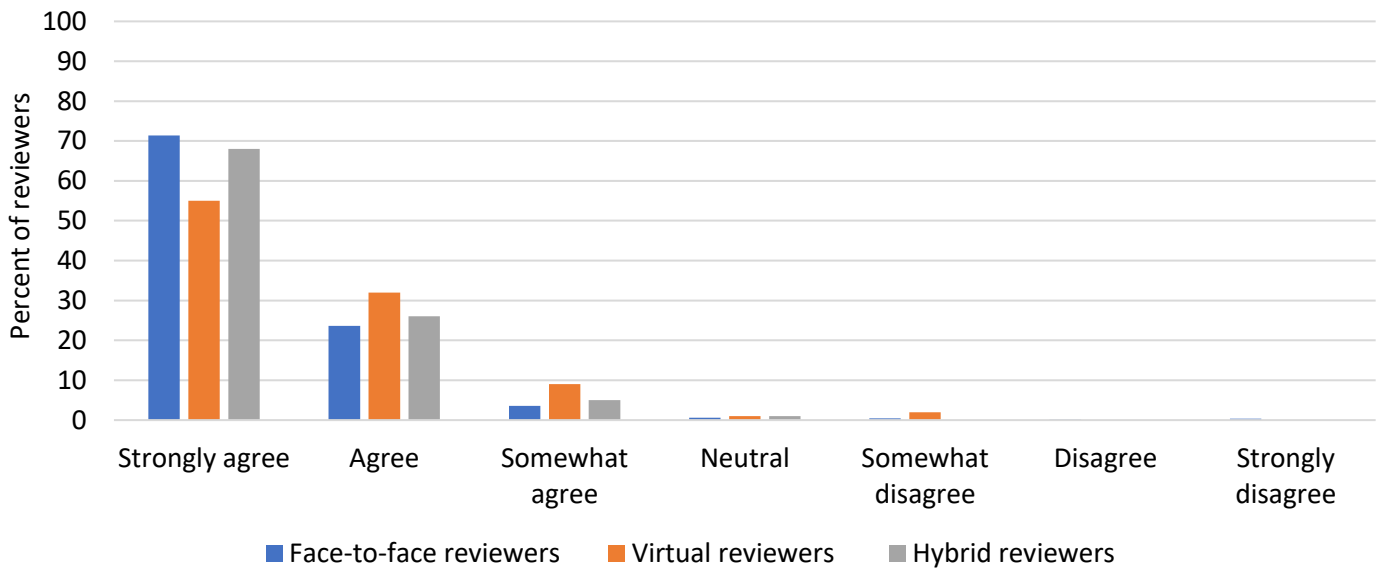
There was a significant effect of meeting format on reviewers' perceptions of whether the discussions helped the panel evaluate the applications, Welch's F(2, 2412.998) = 47.285, $p < .001$, *est. ω^2* = .03. A Games-Howell post hoc test revealed that reviewers who attended face-to-face ($p < .001$, $d = .34$) and hybrid meetings ($p < .001$, $d = .31$) believed that the discussions were significantly more helpful for evaluating the applications than did reviewers who attended virtual meetings. There was no significant difference in reviewers' perceptions on whether the discussions helped the panel evaluate the applications between face-to-face and hybrid meeting formats ($p = .557$).

Survey Data: Examining Hybrid Results According to Mode of Attendance
Examination of the data showed that results for hybrid meetings often appeared to differ according to whether the reviewer attended in person or remotely. Those potential differences are explored below.

Results

The survey was administered to 8,443 reviewers, in which 4,042 completed the survey for a response rate of 48%. Among those that attended hybrid meetings (n = 959), 47% (n = 455) attended remotely and 53% (n = 504) attended in-person[4,5]. See Table 4 for reviewer characteristics. The analyses that follow are confined to hybrid meetings only (n = 959). However, to add context, data from CSR's Summer 2023 survey of face-to-face ("F2F reference data") and virtual ("Virtual reference data") meetings are displayed in many of the figures.

Table 4. Reviewer Characteristics of Survey Respondents for Hybrid Meetings		
Reviewer Characteristics		**% Hybrid Survey Respondents (n = 959)**
Gender		
	Male	52
	Female	45
	Withheld	3
Race		
	American Indian or Alaska Native	< 1
	Asian	20
	Black or African American	4
	More than one race	3
	Native Hawaiian or Pacific Islander	< 1
	White	64
	Withheld	9
Ethnicity		
	Hispanic/Latino	11
	Non-Hispanic	84
	Withheld	5
URM		
	No	77
	Yes	15
	Withheld	8
Career Stage		
	Professor	46
	Associate Professor	29

---

[4] To increase sample size, participants who attended hybrid meetings came from four different rounds of review meetings—155 participants attended the Summer 2023 meetings, 95 participants attended the Fall 2023 meetings, 287 participants attended the Winter 2024 meetings, and 437 participants attended the Summer 2024 meetings, for a total of 974 hybrid responses.
[5] For reviewers who responded to more than one hybrid survey, the first survey response(s) were excluded to avoid non-independent data--resulting in the removal of 15 cases/survey responses (and an adjusted total sample size of 4,027 and hybrid sample size of 959).

	Assistant Professor	18
	Other	7

Meeting Format Preferences for Hybrid Reviewers

- Figure 7 shows the preferences of reviewers from hybrid meetings[6]. Preferences of those reviewers attending face-to-face and virtual review meetings are also shown for comparison.
- Meeting format preference varies based on the format of the meeting attended. In-person reviewers strongly prefer face-to-face meetings, whereas opinions of remote attendees are more divided. Remote attendees in hybrid meetings were especially likely to prefer attending remotely.

Figure 7. Meeting Format Preference for Hybrid Reviewers by Meeting Format
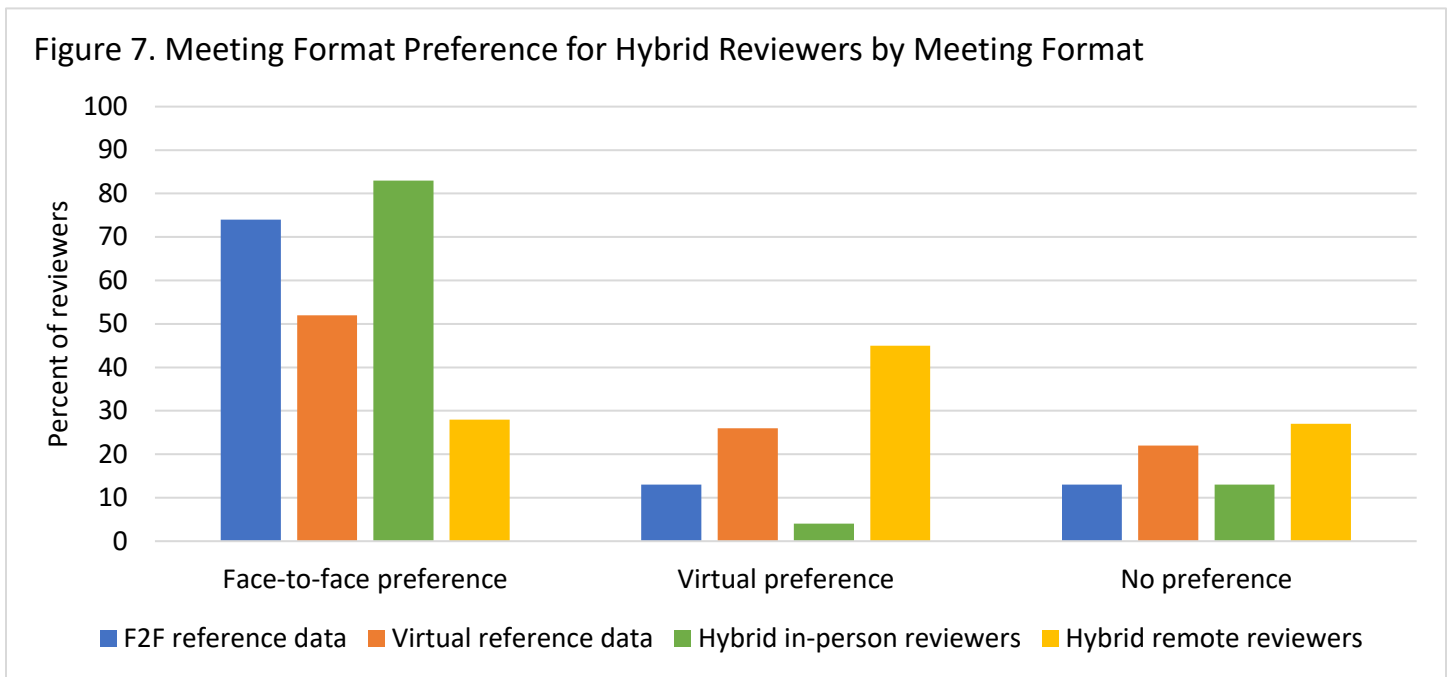
There was a significant association between the format of reviewers' meetings and their meeting format preferences $\chi 2$ (2) = 313.375, p <.001, with a large effect size (ϕ_c = .572, p <.001). Post hoc z-tests show that 1) the proportion of hybrid in-person reviewers that preferred face-to-face meetings was significantly more than the proportion of hybrid remote reviewers that preferred face-to-face meetings, and 2) the proportion of hybrid in-person reviewers that preferred virtual meetings or had no meeting preference was significantly less than the proportion of hybrid remote reviewers that preferred virtual meetings or had no meeting preference.

---

[6] Scientific review officers (SROs) volunteered to hold hybrid meetings, taking into consideration their desired meeting make-up and the preferences of their reviewers on whether they wanted to attend in person or remotely. Thus, these findings could reflect those reviewers who, a priori, preferred to attend in person or remotely.

Quality of Review: Hybrid In-person vs. Hybrid Remote Reviewers

- Figure 8 shows hybrid reviewers' perceptions of the quality of the review meeting. 98% of all hybrid reviewers thought the overall quality of the meetings was good or excellent.

- In-person reviewers rated overall review quality, productivity of discussions, and reviewer engagement better than did remote attendees (see Table 5).

- **Reviewers at face-to-face meetings and hybrid in-person reviewers did not differ on their ratings of overall quality of review or productivity of discussions**. **Ratings of reviewer engagement were better for the reviewers in face-to-face meetings.**

- **In comparison to reviewers at virtual meetings, hybrid remote reviewers rated the quality of the review (across all measures) significantly higher** (see Table 5b).

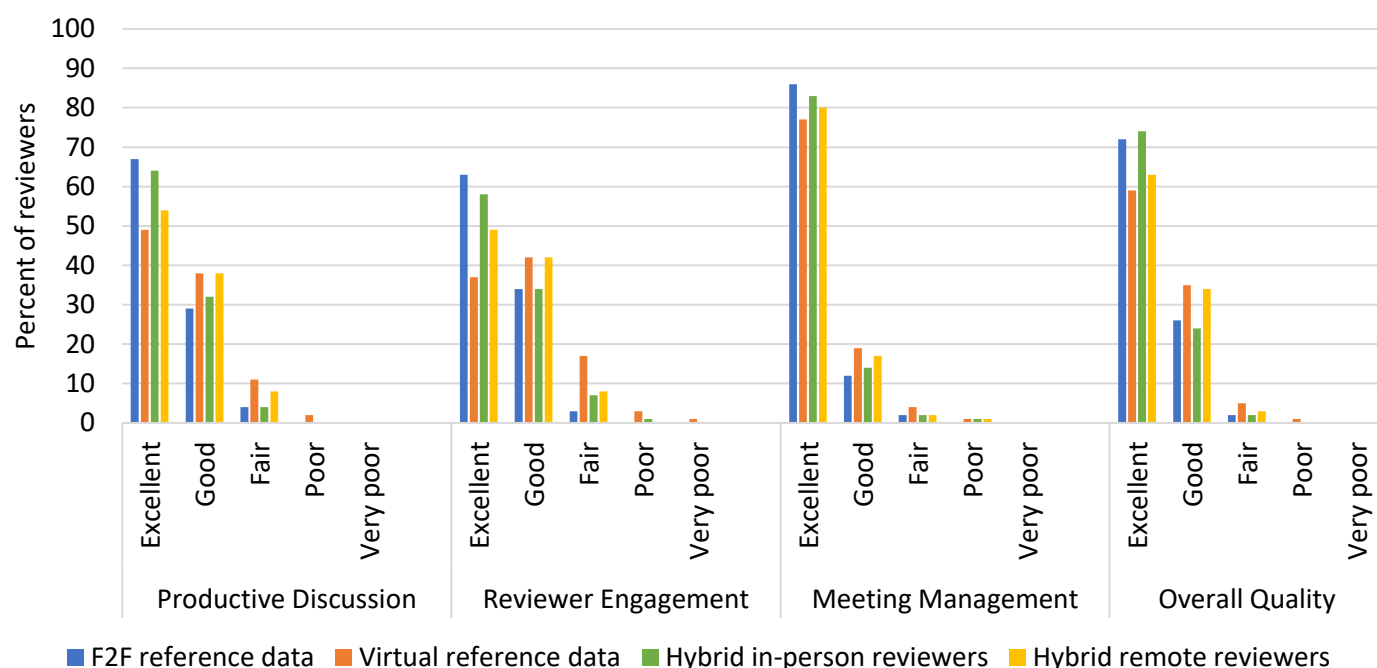Figure 8. Quality of Review for Hybrid Reviewers by Meeting Format

Table 5. Reviewer Ratings of Review Quality of Review by Meeting Format			
	Hybrid In-person Reviewers (M, SE)	Hybrid Remote Reviewers (M, SE)	Independent t-test Statistic and d Between Hybrid Reviewers
Productive Discussion	4.60 (.03)	4.45 (.03)	$t(915.745) = -3.622$, $p < .001$; $d = -.24$
Reviewer Engagement	4.50 (.03)	4.40 (.03)	$t(957) = -2.271$, $p = .023$; $d = -.15$
Meeting Management	4.80 (.02)	4.77 (.02)	$t(924.027) = -1.173$, $p = .241$; $d = -.08$
Overall Quality of Review	4.72 (.02)	4.60 (.03)	$t(912.012) = -3.437$, $p < .001$; $d = -.22$

Table 5b. Ratings of Review Quality: Matched attendee type, different meeting formats						
	Remote attendees in Virtual vs. Hybrid meetings			In-person attendees in Face to Face vs. Hybrid meetings		
	Virtual meeting reviewers (*M, SE*)	Remote hybrid reviewers (*M, SE*)	Independent t-test statistic and *d* between reviewers	Face-to-face meeting reviewers (*M, SE*)	In-person hybrid reviewers (*M, SE*)	Independent t-test statistic and *d* between reviewers
Productive Discussion	4.34 (.02)	4.45 (.03)	$t(818.132) = 3.098$, $p = .005$; $d = .15$	4.61 (.02)	4.60 (.03)	$t(1667) = -.415$, $p = .678$; $d = -.022$
Reviewer Engagement	4.13 (.02)	4.40 (.03)	$t(899.118) = 7.441$, $p < .001$; $d = .34$	4.58 (.02)	4.50 (.03)	$t(858.641) = -2.537$, $p = .011$; $d = -.14$
Meeting Management	4.70 (.02)	4.77 (.02)	$t(809.641) = 2.336$, $p = .020$; $d = .113$	4.84 (.01)	4.80 (.02)	$t(888.874) = -1.261$, $p = .208$; $d = -.07$
Overall Quality of Review	4.53 (.02)	4.60 (.03)	$t(777.900) = 2.270$, $p = .023$; $d = .113$	4.70 (.02)	4.72 (.02)	$t(1667) = .612$, $p = .541$; $d = .033$

Meeting Experience and Participation for Hybrid Meetings

- There were significant differences among hybrid reviewers across all measures of meeting experience, with those who attended in person rating their experience and participation higher (see Table 6). The magnitude of these effects is mostly small. The effect size for reviewers' contribution to the discussion was medium.
- Remote reviewers in hybrid and virtual meetings did not differ in their ratings of meeting experience.
- In comparison to reviewers at face-to-face meetings, hybrid in-person reviewers rated their meeting experience and participation higher on all measures (see Table 6b)--with their contribution to the discussion and confidence voicing their opinions significantly higher.

Figure 9. Experience and Participation for Hybrid Reviewers by Meeting Format
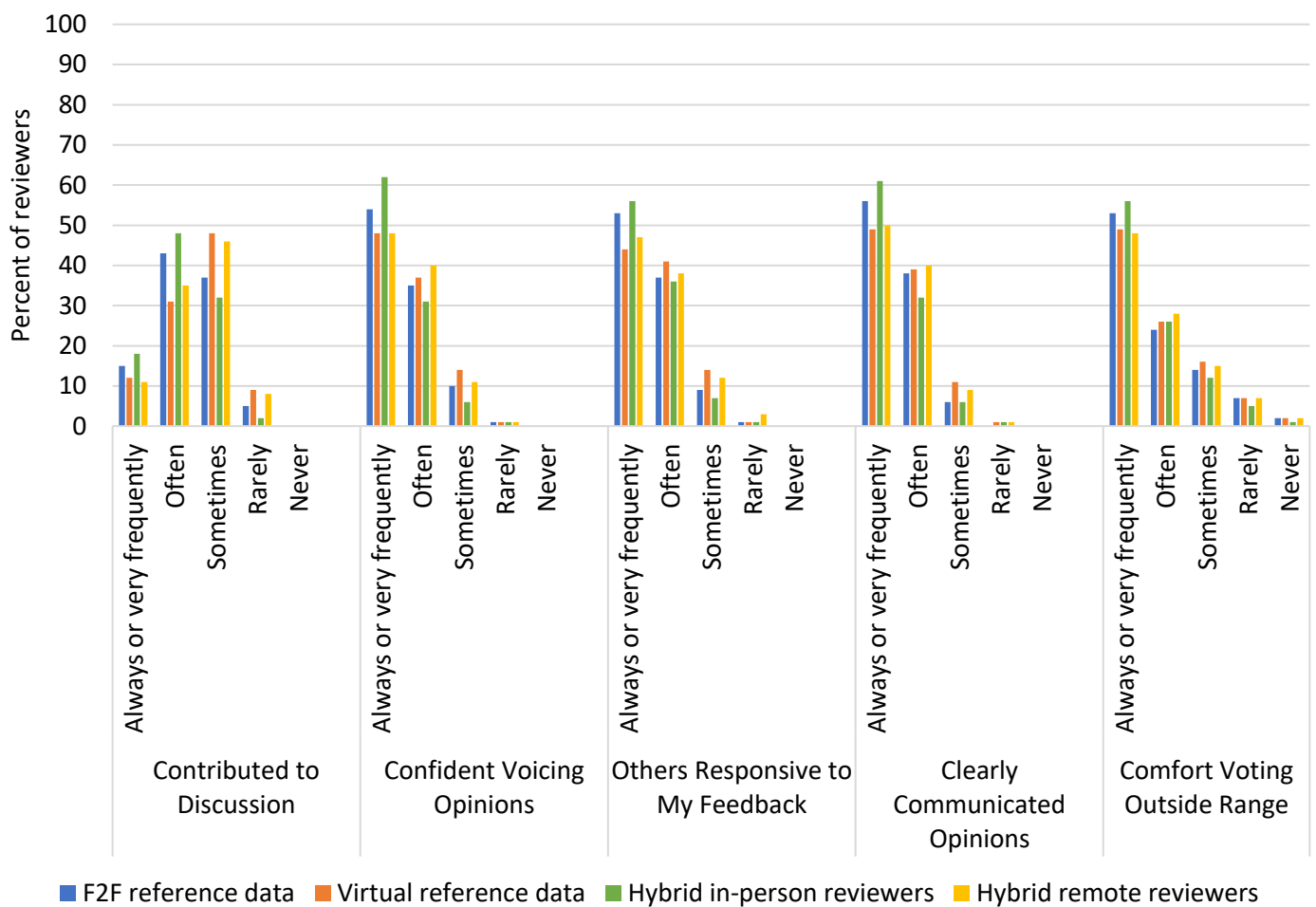
Table 6. Experience and Participation of Reviewers Attending Hybrid Meetings

	Hybrid In-person Reviewers (*M, SE*)	Hybrid Remote Reviewers (*M, SE*)	Independent t-test Statistic and *d* Between Hybrid Reviewers
Contributed to Discussion	3.82 (.03)	3.49 (.04)	$t(928.315) = -6.497$, $p < .001$; $d = -.40$
Confident Voicing Opinions	4.53 (.03)	4.34 (.03)	$t(926.480) = -4.212$, $p < .001$; $d = -.27$
Others Receptive and Responsive to Feedback	4.47 (.03)	4.31 (.04)	$t(872.566) = -3.433$, $p < .001$; $d = -.23$
Clearly Communicated Opinions	4.53 (.03)	4.40 (.03)	$t(955) = -3.115$, $p = .002$; $d = -.20$
Comfortable Voting Outside Range	4.29 (.05)	4.13 (.05)	$t(836) = -2.350$, $p = .019$; $d = -.16$

Table 6b. Ratings of Review Experience and Participation: Matched attendee type, different meeting formats

	Remote attendees in Virtual vs. Hybrid meetings			In-person attendees in Face to Face vs. Hybrid meetings		
	Virtual meeting reviewers (*M, SE*)	Remote hybrid reviewers (*M, SE*)	Independent t-test statistic and *d* between reviewers	Face-to-face meeting reviewers (*M, SE*)	In-person hybrid reviewers (*M, SE*)	Independent t-test statistic and d between reviewers
Contributed to Discussion	3.46 (.02)	3.49 (.04)	$t(2157) = .851$, $p = .395$; $d = .045$	3.69 (.02)	3.82 (.03)	$t(1021.354) = 3.169$, $p = .002$; $d = .16$
Confident Voicing Opinions	4.31 (.02)	4.34 (.03)	$t(2152) = .812$, $p = .417$; $d = .043$	4.41 (.02)	4.53 (.03)	$t(1042.700) = 3.485$, $p < .001$; $d = .18$
Others Receptive and Responsive to Feedback	4.27 (.02)	4.31 (.04)	$t(2127) = .857$, $p = .392$; $d = .046$	4.43 (.02)	4.47 (.03)	$t(1653) = 1.192$, $p = .234$; $d = .06$
Clearly Communicated Opinions	4.36 (.02)	4.40 (.03)	$t(2150) = .926$, $p = .354$; $d = .049$	4.50 (.02)	4.53 (.03)	$t(1658) = 1.057$, $p = .291$; $d = .06$
Comfortable Voting Outside Range	4.13 (.03)	4.13 (.05)	$t(1805) = -.029$, $p = .977$; $d = -.002$	4.20(.03)	4.29 (.05)	$t(1448) = 1.576$, $p = .115$; $d = .09$

Reviewers' Attention Span in Hybrid Meetings

- Figure 10 shows hybrid reviewers' report of their ability to pay attention throughout the meeting. Hybrid reviewers who attended in person rated their attention better than did those who attended remotely. The magnitude of this effect is small.
- In comparison to face-to-face meetings, in-person hybrid reviewers rated their attention span about the same[7].
- **In comparison to virtual meeting attendees, remote hybrid reviewers rated their attention span significantly better[8].**

Figure 10. Reviewers' Attention Span for Hybrid Reviewers by Meeting Format
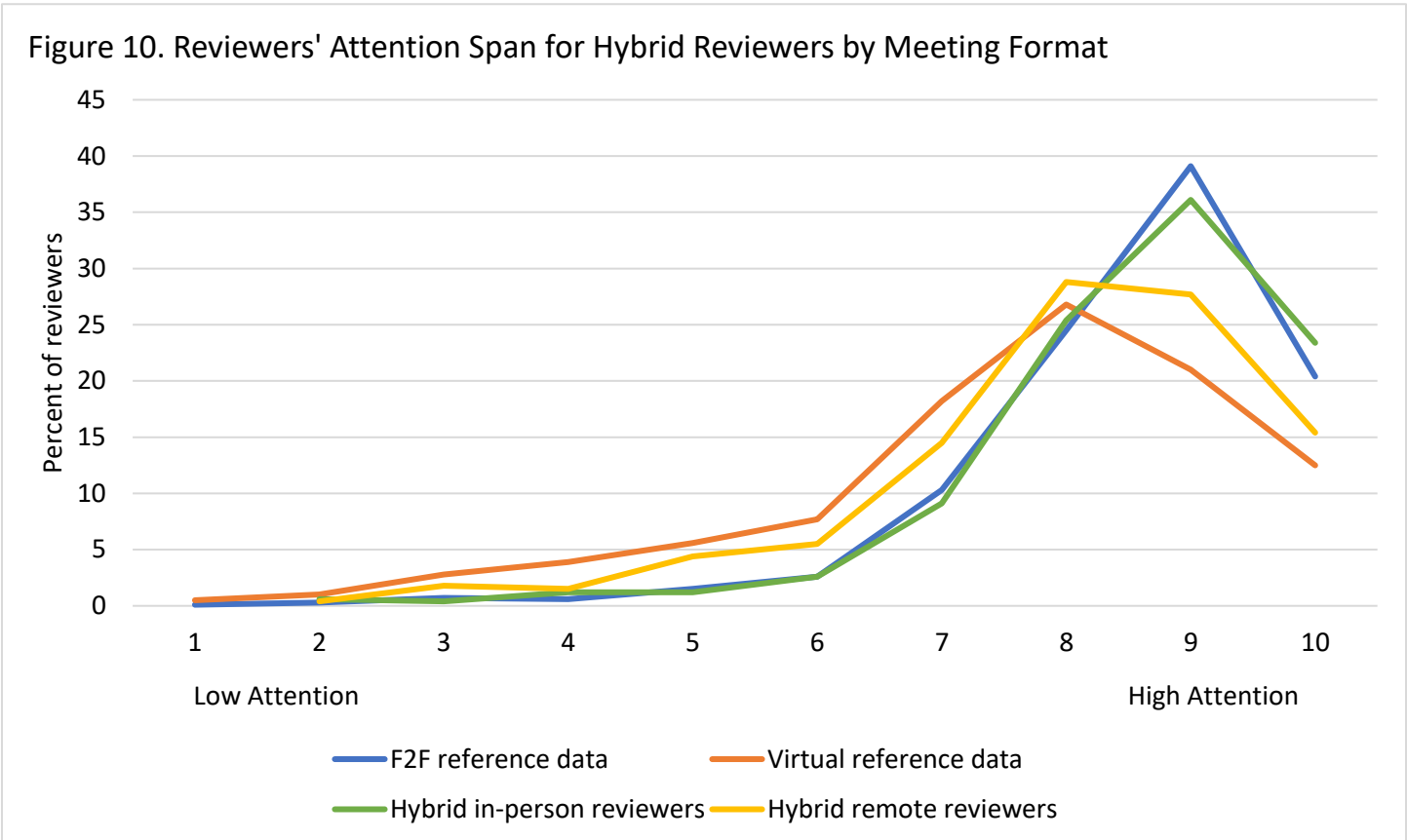
There was a significant difference between hybrid reviewers who attended the meeting in-person (M = 8.55, SE = .06) and those who attended remotely (M = 8.02, SE = .07) in their attention span or ability to concentrate at the meeting. This difference, -.526, CI [-.713, -.339], was significant $t(895.451)$ = -5.515, $p < .001$; and represented a small-sized effect, d = -.36
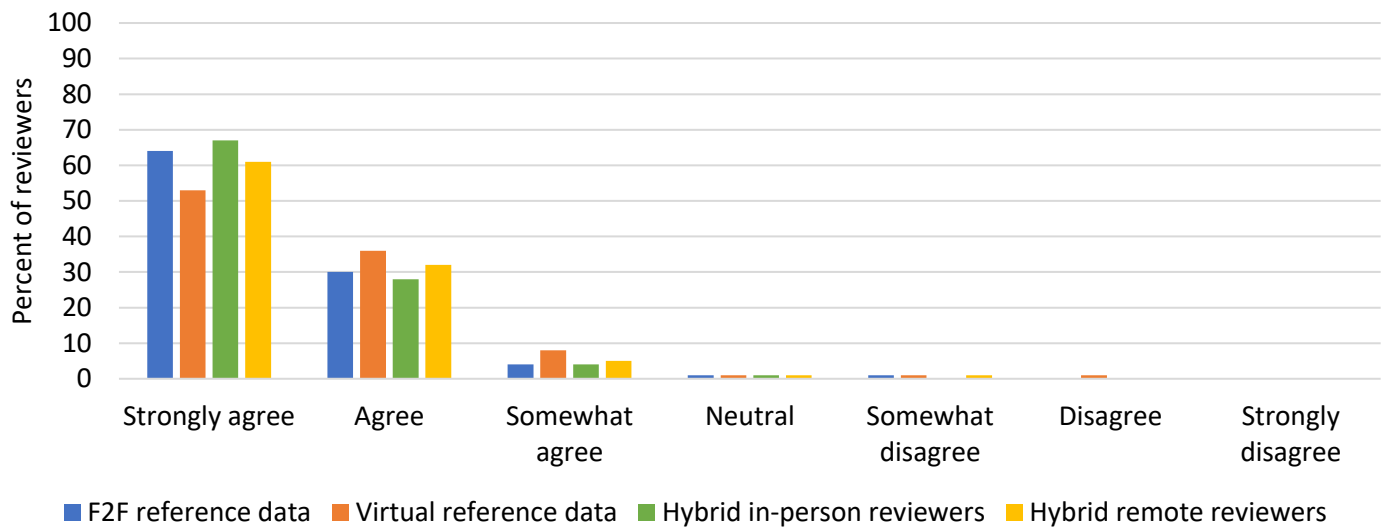
---

[7] F2F reference data, M = 8.52, SE = .04; $t(1669)$ = .452, p = .651 (for comparison with in-person hybrid reviewers).
[8] Virtual reference data, M = 7.56, SE = .05; $t(819.780)$ = 5.286, $p < .001$, d = .25 (for comparison with remote hybrid reviewers).

rioritizing Applications

- Figure 11 shows data capturing hybrid reviewers' perceptions of the panel's ability to prioritize applications. Overall, 94% of all hybrid reviewers believed that the panel was able to prioritize applications according to their impact and scientific merit.
- There was no significant difference in rated ability to prioritize applications between those remote and in-person attendees.
- **In comparison to face-to-face meeting attendees, in-person hybrid reviewers rate their ability to prioritize applications about the same[9].**
- **In comparison to virtual meetings, remote hybrid reviewers rated their ability to prioritize applications significantly better[10].**

Figure 11. Assessment of the Panel's Ability to Prioritize Applications for Hybrid Reviewers by Meeting Format

Note: Survey questions were on a scale from 1 (strongly agree) to 7 (strongly disagree), with lower scores indicating a more favorable view of the review meeting.

The ability to prioritize applications between those hybrid reviewers who attended the meeting in person (M = 1.40, SE = .03) and those who attended remotely (M = 1.49, SE = .04) was not significantly different (.088, CI [-.004, .179], t(917.894) = 1.883, p = .060).

---

[9] F2F reference data, M = 1.46, SE = .02; t(1669) = -1.358, p = .175 (for comparison with in-person hybrid reviewers).
[10] Virtual reference data, M = 1.64, SE = .02; t(793.100) = - 3.516, p < .001, d = -.17 (for comparison with remote hybrid reviewers).

Perceptions of Discussion Quality

- Figure 12 shows data capturing hybrid reviewers' perceptions of the panel's ability to discuss and evaluate applications. Overall, 94% of all hybrid reviewers believed that the scientific discussions helped the panel evaluate the applications being reviewed.
- There was a significant difference between hybrid reviewers who attended in person and those who attended remotely, with hybrid reviewers who attended in person reporting that the discussions were more helpful for evaluating the applications.
- In comparison to face-to-face meeting attendees, in-person hybrid reviewers rated the utility of discussions in evaluating applications about the same[11].
- **In comparison to virtual meetings, remote hybrid reviewers rated the utility of discussions in evaluating applications significantly higher or more useful[12].**

Figure 12. Discussions Helped with Evaluations for Hybrid Reviewers by Meeting Format

Note: Survey questions were on a scale from 1 (strongly agree) to 7 (strongly disagree), with lower scores indicating a more favorable view of the review meeting
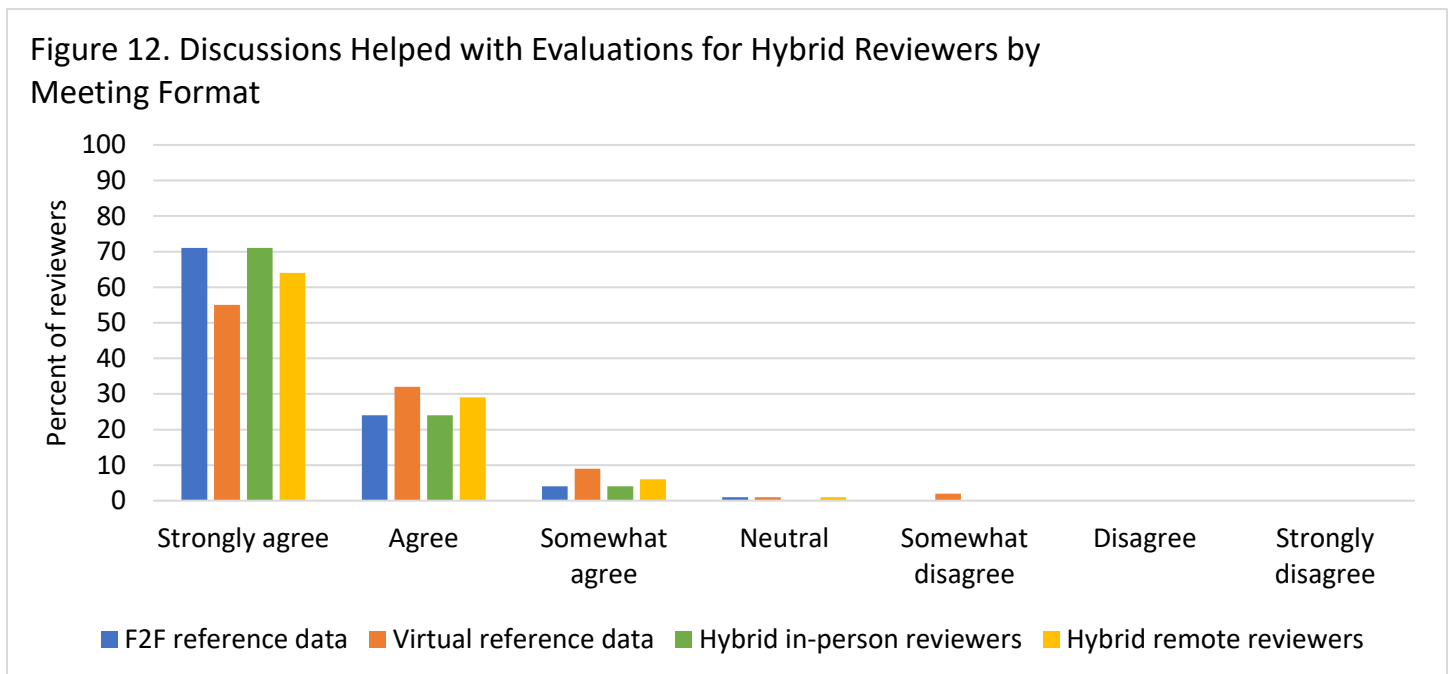
There was a significant difference between hybrid reviewers who attended the meeting in person (M = 1.36, SE = .03) and those who attended remotely (M = 1.45, SE = .03) in their perceptions of whether the discussions helped the panel evaluate the applications. This difference, .091, CI [.004, .178], was significant $t(933.320)$ = 2.063, p = .039, and represented a small-sized effect, d = .13.

---

[11] F2F reference data, M = 1.37, SE = .02; t(1669) = -.318, p = .750 (for comparison with in-person hybrid reviewers).
[12] Virtual reference data, M = 1.65, SE = .02; t(904.048) = - 4.944, p < .001, d = -.23 (for comparison with remote hybrid reviewers).

21

Quantitative Meeting Measures – Roster Composition and Scores
See *Appendix C* for detailed methods.

Quantitative Meeting Measures Highlights
- There were no meaningful differences between face-to-face, virtual, and hybrid meetings in terms of the diversity (e.g. career stage, part of the country, demographics) of the ad hoc reviewers recruited.

- Meeting format was associated with small differences in scoring behavior. Final overall impact scores were slightly worse in face-to-face and hybrid meetings compared to those in virtual meetings. The effect size was small. The mean score change was larger in face-to-face and hybrid meetings than in virtual meetings. The effect size was small. There were more out-of-range scores in face-to-face and hybrid meetings than in virtual meetings – the effect size is small.

Methods
Roster characteristics and scoring practices were compared for the three meeting formats: face-to-face, virtual, and hybrid meetings. Data on rosters and scores for all standing study sections, fellowships, and small business panels (n=227) for face-to-face and virtual meetings that occurred in Summer 2023 were collected. To maximize the amount of data on hybrid meetings, rosters and scores for all standing study sections, fellowships, and small business panels that met as hybrids in Summer 2023, Fall 2023, Winter 2024, and Summer 2024 (n=62) were collected (Table 7). Detailed methods can be found in [Appendix C](#).

Table 7. Number of Meetings Included in the Analysis		
Meeting format	**n of meetings**	**Council rounds included in the analysis**
Face-to-face Meetings	82	Summer 2023
Virtual Meetings	145	Summer 2023
Hybrid Meetings	62	Summer 2023, Fall 2023, Winter 2024, Summer 2024
Total	**289**	

Meeting Application Counts, Roster Sizes, Reviewer Workload Trends
Meeting Application Counts
The average application count (Table 8) for face-to-face, virtual, and hybrid meetings was 70, 68, and 73, respectively. An ANOVA test revealed that there is statistical significance in mean application counts by meeting format. Howell's HSD test found that the mean value of application counts between hybrid and virtual meetings was significantly different (5.545, *CI* [.243, 10.846]). There were no statistically significant differences in mean application counts between face-to-face and virtual meetings or between face-to-face and hybrid meetings.

Table 8. Mean Application Counts				
	Face-to-face Meetings (*M*, *SE*)	**Virtual Meetings (*M*, *SE*)**	**Hybrid Meetings (*M*, *SE*)**	**ANOVA**
Application Counts	70.02 (1.64)	67.46 (1.22)	73.00 (1.88)	$F_{(2,286)}=3.15$, *p*=.045

Roster Size and Reviewer Workloads

Face-to-face, virtual, and hybrid meetings did not differ in terms of the number of reviewers recruited (roster size) or in the average number of applications assigned to each reviewer (Table 9).

Table 9. Roster Size and Reviewer Workload				
	Face-to-face Meetings (*M*, *SE*)	Virtual Meetings (*M*, *SE*)	Hybrid Meetings (*M*, *SE*)	ANOVA
Roster Size	29.05 (0.63)	28.38 (0.47)	30.23 (0.72)	F(2,286)=2.33, *p*=.100
Reviewer Workload	7.66 (0.11)	7.46 (0.08)	7.67 (0.12)	F(2,286)=1.55, *p*=.213

Roster Composition – characteristics of ad hoc reviewers

Effects of meeting format on roster composition were assessed. Only ad hoc members were analyzed, as standing members have committed to service throughout their term, and whether a meeting is held in-person or virtually in any given round is unlikely to influence their meeting attendance. Focusing on ad hoc members allows the analyses to determine the effect of meeting formats on reviewer recruitment in relation to their seniority, geographic location, demographics, and extent of prior service (focusing on new reviewers).

Faculty Rank

It has been hypothesized that more senior reviewers may prefer face-to-face meetings than do reviewers in earlier career stages. Table 10 displays the distribution of faculty rank for ad hoc members by meeting format. A Pearson chi-square test revealed no statistically significant differences in the distribution of faculty rank between the three meeting formats, $\chi2 (6) = 8.986$, $p = .174$.

Table 10. Distribution of Academic Title for Ad Hocs			
	Face-to-face Meetings	Virtual Meetings	Hybrid Meetings
Professor	42.21%	39.95%	39.86%
Associate Professor	34.15%	34.00%	33.14%
Assistant Professor	16.13%	17.79%	17.77%
Other	7.52%	8.26%	9.23%

Geographic Location

As virtual and hybrid meeting formats remove travel as a participation barrier, it has been hypothesized that those who reside far from the meeting location may be more likely to participate in review for virtual or hybrid meetings. At the time of this report, all face-to-face and hybrid meetings are held in the Washington, D.C. area, so one may expect to see a higher proportion of reviewers from the West or Central regions of the U.S. in virtual or hybrid meetings than face-to-face meetings. Table 11 displays the distribution of geographic location for ad hoc members by meeting format. Reviewers from foreign countries and US territories were excluded from this analysis (n=10).

A Pearson chi-square test revealed statistically significant differences in geographic distributions of reviewers between the three meeting formats, $\chi2 (6) = 16.621$, $p = .011$, with a small effect size ($\phi_c = 0.03$). Upon further investigation of standardized residuals, "South, Hybrid" and "East, Hybrid" have large standardized residuals. Observed ad hoc reviewers in "South, Hybrid" were lower than expected, and observed ad hoc reviewers in "East, Hybrid" were higher than expected.

Table 11. Distribution of Geographic Location for Ad Hocs			
	Face-to-face Meetings	Virtual Meetings	Hybrid Meetings
South	31.33%	31.26%	27.19%*
East	25.16%	25.22%	28.96%*

Central	23.72%	23.26%	24.18%
West	19.79%	20.25%	19.67%

Demographics

It has been hypothesized that alternative review formats, such as virtual or hybrid meetings, may have positive effects on the inclusion of women and underrepresented minorities (URMs) in peer review. Table 12 displays the inclusion of URM ad hoc members by meeting format. Those ad hoc members whose URM status is unknown (n=460) were excluded from these analyses. URMs are defined as individuals that identify as Black or African American, Hispanic or Latino, American Indian or Alaska Native, Native Hawaiian, and other Pacific Islanders. A Pearson chi-square test revealed no statistical differences for the inclusion of URM between the three meeting formats (χ2 (2) = 0.599, *p* = .741).

Table 12. Average Ad Hoc URM representation		
Face-to-face Meetings	**Virtual Meetings**	**Hybrid Meetings**
16.18%	15.70%	16.46%

Table 13 displays the inclusion of women ad hoc members by meeting format. Those ad hoc members whose gender status was not reported (n=126) were excluded from these analyses. The distribution of gender between the three meeting formats was not found to be statistically significant, χ2 (2) = 0.216, *p* = .898.

Table 13. Average Ad Hoc Women representation		
Face-to-face Meetings	**Virtual Meetings**	**Hybrid Meetings**
43.86%	44.32%	44.54%

New Reviewer Recruitment

It has been hypothesized that virtual or hybrid meeting formats may remove travel as a participation barrier, thereby incentivizing new reviewers to engage in the peer review process who were unable, opposed to the travel, or ambivalent about participating in face-to-face meetings. Tables 14 and 15 display the average number of ad-hoc reviewers with no to little (1-2 review instances prior meetings) prior review service. An ANOVA test did not reveal any statistical differences in the number of ad hoc reviewers with little to no prior review service participating between the meeting formats.

It should be noted that while standing study sections have standing members that are excluded from this analysis, small business panels and fellowship panels do not have standing members, so the entire panel of ad hoc reviewers is accounted for these meetings. Another policy difference is that CSR requires that two Early Career Reviewers (ECRs; those enrolled in the CSR Early Career Reviewer Program) be recruited for each meeting of a standing panel; ECR recruitment is not mandated for small business or fellowship panels. ECRs were included in this analysis.

Table 14. New Reviewer Recruitment (0 prior reviews)			
Face-to-face Meetings (*M*, *SE*)	**Virtual Meetings *M* (SE)**	**Hybrid Meetings (*M*, *SE*)**	**ANOVA**
3.33 (0.21)	3.21 (0.16)	3.49 (0.24)	F(2,289) = 0.47, *p*=.625

| Table 15. New Reviewer Recruitment (1-2 prior reviews)			
Face-to-face Meetings (*M*, *SE*)	**Virtual Meetings (*M*, *SE*)**	**Hybrid Meetings *M* (SE)**	**ANOVA**
3.34 (0.26)	3.71 (0.20)	3.92 (0.30)	F(2,289) = 1.16, *p*=.314

Scores

It has been hypothesized that review quality may suffer in virtual or hybrid meeting formats – reviewers (especially unassigned reviewers) may be less engaged or distracted in their remote environments, reviewers may be affected by Zoom fatigue, or there may be decreased rapport between panel members and less engaged open discussions. Scoring behaviors between meeting formats as an indicator of review quality, engagement, and panel function were examined.

Final Overall Impact Scores

Table 16 shows the mean value of final overall impact scores by meeting format, and Figure 13 displays the distribution of final overall impact scores of discussed applications only by meeting format. An ANOVA test revealed statistical significance in the mean value of final overall impact scores between the three meeting formats. Tukey's HSD test revealed that the mean value of final overall impact scores had a significant difference between hybrid and virtual meetings (1.290, *CI* [.620, 1.961]) and between face-to-face and virtual meetings (0.901, *CI* [.285, 1.517]). There were no statistically significant differences in mean final overall impact scores between face-to-face and hybrid meetings.

Table 16. Final Overall Impact Scores, Discussed Applications Only			
Face-to-face Meetings *M [SD]*	**Virtual Meetings** *M [SD]*	**Hybrid Meetings** *M [SD]*	**ANOVA**
38.02 [12.0]	37.12 [11.5]	38.82 [12.0]	$F_{(2, 10889)} = 12.19$, *p*<.000

Figure 13. Distribution of final overall impact scores of discussed applications only, by meeting format
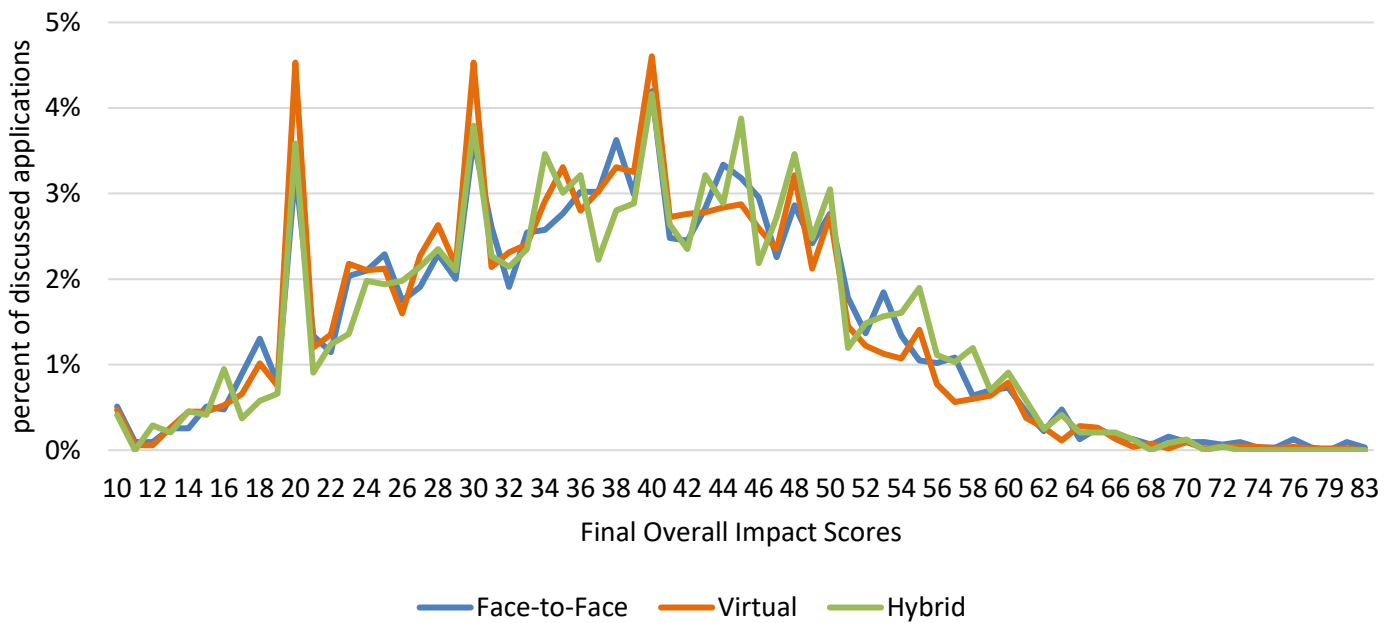
Score Movement: Assigned Reviewers Only

It has been hypothesized that in virtual or hybrid meeting formats, fewer assigned reviewers may change their scores after discussions, as there is a perception that some reviewers are less engaged. The absolute value of score movement from preliminary to final scores assigned to reviewers was evaluated by meeting format (Table 17). Welch's ANOVA test revealed that there was a statistically significant difference in score changes between the meeting formats. An independent t-test for all possible comparisons revealed that differences in score movement from preliminary to final overall impact scores were statistically significant between hybrid and virtual meetings (t(22149)=3.71, p>0.000) and between hybrid and face-to-face meetings (t(20116)=2.58, p>0.010). There were no significant differences between face-to-face and virtual meetings.

Table 17. The Mean and Standard Error of the Absolute Value of Changes of Assigned Reviewers' Preliminary Scores to Final Scores			
Face-to-face Meetings *M (SE)*	Virtual Meetings *M (SE)*	Hybrid Meetings *M (SE)*	Welch's ANOVA
0.46 (0.01)	0.45 (0.01)	0.49 (0.01)	F(2, 21502)= 7.07, p=.001

Out of Range Scores: Unassigned reviewers Only

It has been hypothesized that fewer unassigned reviewers would vote outside of range in virtual or hybrid formats, as they may not be engaged in discussion to elucidate differences of opinion. Table 18 shows the percentage of final scores that were out of range by unassigned reviewers, by meeting format. A Pearson chi-square test revealed that there was a statistically significant association between the meeting format ($\chi2$ (2)=2) =287.07, p<0.000) with a very small effect size (ϕ_c=0.03).

Table 18. Extent of Out-of-Range Scores		
Face-to-face Meetings	Virtual Meetings	Hybrid Meetings
4.1%	3.3%	4.7%

Scores: Hybrid Meetings Only

This section examined the hybrid meeting only, specifically comparing scores between in-person and remote reviewers. A total of 62 hybrid meetings from the Summer 2023, Fall 2023, Winter 2024, and Summer 2024 council rounds were included, and a total of 1,874 reviewers participated in these hybrid meetings. Approximately 46% of the hybrid meeting reviewers attended in person, and the other 54% attended remotely (Table 19).

Table 19. Distributions of Hybrid Meeting Reviewers	
Attendance Type	**n of reviewers**
Hybrid in-person reviewers	859
Hybrid remote reviewers	1015
Total	**1874**

The absolute values of score movement from preliminary to final scores by assigned reviewers were examined (Table 20). A Satterthwaite t-test was performed to examine the difference in score movement between hybrid in-person and hybrid remote reviewers, and was found to be statistically significant, with a small effect size.

Table 20. Hybrid Only: The Mean and Standard Error of the Absolute Value of Changes of Assigned Reviewers' Preliminary Scores to Final Scores		
Hybrid in-person reviewers *M (SE)*	**Hybrid remote reviewers** *M (SE)*	**t-test**
0.54 (0.01)	0.47 (0.01)	$t(10471)=3.49$, $p=.001$, $d=0.07$

Out-of-range scores between hybrid in-person and hybrid remote reviewers by unassigned reviewers were evaluated. Table 21 shows the percentage of final scores that were out of range by unassigned reviewers, by meeting attendee types. A Pearson chi-square test showed that there was a significant association in out-of-range scores between hybrid in-person and hybrid remote reviewers, $\chi 2\ (1) = 48.39$, $p < .000$, with a small effect size (ϕ_c =-0.02).

Table 21. Hybrid Only: Extent of Out-of-Range Scores	
Hybrid In-person Reviewers	**Hybrid Remote Reviewers**
5.2%	4.3%

Appendix A

Table 22. Information on Meeting Formats, Reviewer Attendance, and Data included in analyses

Meeting Format	Face-to-Face Meetings	Virtual Meetings	Hybrid Meetings
How reviewers attend	Face-to-face reviewers only	Virtual reviewers only	Hybrid reviewers (Sub analyses: hybrid in-person reviewers vs hybrid remote reviewers)
Meeting timeframe	Summer 2023 meetings only	Summer 2023 meetings only	Summer 2023, Fall 2023, Winter 2024, Summer 2024
Meeting council round	2023/08 or 2023/10	2023/08 or 2023/10	2023/08, 2023/10, 2024/01, 2024/05, 2024/08, 2024/10
Included in analyses			
Surveys: between meeting format	✔	✔	✔
Surveys: hybrid sub analyses	✔ (as a comparison point)		✔
Quantitative meeting measures: Meeting application counts, roster sizes, reviewer workload trends	✔	✔	✔
Quantitative meeting measures: Roster composition – characteristics of ad hoc reviewers	✔	✔	✔
Quantitative meeting measures: scores between meeting formats	✔	✔	✔
Quantitative meeting measures: scores between attendee types in hybrid meetings			✔

Appendix B
Survey Methods

Participants

Reviewers who participated in 291 CSR study section meetings (n = 4,042) between May 31, 2023, and July 25, 2024. The study section meetings included chartered panels, recurring Small Business and Fellowship special emphasis panels (SEPs), and a few other SEPs that held hybrid review meetings. Reviewers who attended hybrid meetings came from 60 study sections.

Survey Administration

Reviewers were asked for their participation in a survey via email on the last day of the study section meeting. The email contained a weblink to the survey. Reviewers were told in the email that their responses would be kept confidential and that the survey would take less than five minutes to complete. All surveys returned by September 13th, 2024, were included for analysis.

Measures

Application Evaluation

Two items asked participants to rate on a scale from 1 (strongly agree) to 7 (strongly disagree) the panel's ability to evaluate the applications: 1) the panel was able to prioritize applications according to their impact and scientific merit, and 2) the scientific discussion helped the panel evaluate the applications being reviewed.

Peer Review Quality

Four items asked participants to rate on a scale from 1 (very poor) to 5 (excellent) the following items: 1) overall quality of review, 2) productivity of discussions, 3) level of reviewer engagement, and 4) meeting management.

Reviewer Meeting Experience and Participation

Five items asked participants to rate on a scale from 1 (never) to 5 (always or very frequently) the following items: 1) I contributed to the discussion, 2) I felt confident voicing my opinions, 3) I felt others were receptive and responsive to my feedback, 4) I was able to clearly communicate opinions, and 5) I felt comfortable voting outside the range.

Attention Span

One item asked participants to rate their attention span at the review meeting using a scale from 1-10, with 1 being really struggling to concentrate and 10 being no problem concentrating at all.

Format Preference

One question asked participants if there were no or minimal health risks from COVID-19, would they be more likely to participate in a review meeting if it was held face to face or over Zoom/video? Response options included: face-to face, Zoom/video, and no preference.

Demographic Information

Four questions were used to collect the demographic characteristics of respondents. 1) *Gender*: male, female, I prefer not to respond; 2) *Race:* American Indian or Alaskan Native, Asian, Black or African American, Native Hawaiian or

Pacific Islander, White, I prefer not to respond; 3) *Ethnicity*: Are you Hispanic? Yes, No, I prefer not to respond; 4) *Career stage*: Assistant Professor, Associate Professor, Professor, Other.

Participants' race and ethnicity were used to determine whether they were an underrepresented minority or not. Non-Hispanic Asians and Non-Hispanic Whites were coded as "not URM," and all other participants were coded as "URM." For participants who identified with more than one racial group, if one racial identity was not White or Asian, they were coded as "URM." Participants who identified as both White and Asian were coded as "not URM." According to OMB standards, individuals who identify with an Asian racial group other than Chinese, Filipino, Japanese, Korean, Asian Indian, or Thai are considered an under-represented minority (URM). However, the current data does not allow for this level of group specificity, and therefore only Asian participants who identify as Asian and another racial group (other than White) or as Asian and Hispanic are coded as URM.

Open-ended Response Options

In an open-ended text box, participants were asked to please share any comments (positive or negative) about their experience or general thoughts on their recent review meeting.

Appendix C

Methods for Quantitative Meeting Measures
General meeting information was extracted from CSR's internal Meeting Dashboard. Roster data (such as information on titles (rank) of the reviewer, review expertise, etc.) was extracted from the internal NIH IMPAC II database (QVR). Sensitive data such as demographics and scores were provided by the CSR Informatics team. Mail reviewers were excluded for all analyses, while ECRs were included but not specified.

For roster analyses by gender or URM, those ad hoc members whose gender status was unknown (n=126) or URM status was unknown (n=460) were excluded from these analyses.

Extent of prior review service was measured for ad hoc reviewers only participating in one of the meetings in the sample. CSR Informatic team provided cross-sectional reviewer-level data on extent of prior review, adjusted for one week before the meeting to get the most accurate data at the time of recruitment. Meeting counts for prior extent of service span a 12-year period and included NIH review meetings (both those run by CSR and by other NIH institutes/centers) and Advisory Councils for NIH institutes/centers. Meetings not classified as a meeting of a Federal Advisory Committee, mail reviews, and CSR rump Special Emphasis Panels were excluded from meeting counts.

Limitation
ECRs were not separated out, so all roster analyses metrics will include their metrics as well, including prior review service, where they will disproportionately represent reviewers in standing study sections who have had 0 prior meetings and are "newly engaged in the review system."

Reviewer Survey Highlights: Hybrid Meetings Only
The data shows that hybrid reviewers collectively perceived hybrid meetings as highly effective; over 90% of reviewers rated discussions, reviewer engagement, and overall meeting quality as excellent or good. However, significant differences with small effect sizes were observed between in-person and remote reviewers' ratings of meeting quality and personal experience of the meetings. Reviewers' preferences for meeting format differed substantially depending on whether they attended in person or remotely.

- 94% of reviewers in hybrid meetings thought that their review panels were able to prioritize applications according to their impact and scientific merit and that the scientific discussions helped the panel evaluate the applications. There were no significant differences between reviewers' ratings on prioritizing applications; reviewers who attended hybrid meetings in person rated the discussions as significantly more helpful for evaluating the applications, although the magnitude of this effect is small.

- 98% of reviewers in hybrid meetings rated the overall quality of review at their meetings as excellent or good. Reviewers who attended hybrid meetings in person rated the quality of their reviews significantly higher (across all measures, except for meeting management) than reviewers who attended remotely, although the magnitude of these effects is small.

- Reviewers reported high levels of self-participation and a positive experience at their hybrid review meetings. Ratings of participation and engagement were significantly higher from reviewers who attended in person vs. remotely, although the magnitude of these effects is small.

- Reviewers who attended hybrid meetings in person reported being able to sustain attention longer than did remote reviewers.

- A majority of hybrid reviewers prefer meeting in person, but results differ greatly according to how they attended. In-person reviewers favor face-to-face meetings by 83% to 4%, while remote reviewers favor virtual meetings by 45% to 28%.