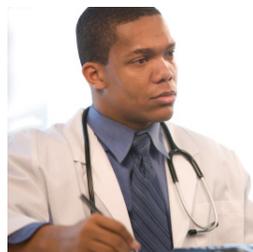


Enhancing Peer Review Survey Results Report



Published May 2013



Executive Summary

The Phase II Enhancing Peer Review surveys, conducted in spring 2012, elicited opinions about the NIH peer review system from recent NIH grant applicants, reviewers, NIH advisory council members, Scientific Review Officers (SROs) and Program Officials (POs). The survey questions focused on changes introduced after the Phase I surveys were completed in Spring 2010, including the shortened, realigned grant application, the single resubmission policy, and the narrative Overall Impact statement. The Phase II surveys also assessed whether stakeholders' opinions had changed since the Phase I surveys were conducted, focusing on areas that represent core peer review performance benchmarks, and areas identified in Phase I as needing further attention.

The findings showed that most applicants agreed that most sections of the shortened applications are sufficient to describe their research project; applicants agreed less often that the Introduction to Resubmission section was sufficient. Reviewers, POs and SROs agreed that most shortened application sections were sufficient to evaluate the scientific merit of most research grant applications. However, rates of agreement were lower among all three stakeholder groups that the appendix section was sufficient. Program Officers agreed less often that the Research Strategy section was sufficient for applications proposing clinical research projects.

Although most reviewers agreed that the nine-point scoring scale has sufficient range to evaluate most grant applications, their open-ended comments — which were more consistent with responses from other groups — indicated concern about reviewers' tendency to assign scores unevenly across the range of available scores. Reviewers expressed a desire for more scoring guidance, coaching and direct scoring instruction.

SROs, POs and advisory council members disagreed more often that bulleted comments were composed in complete sentences or complete thoughts. However, they agreed that the summary statement content is overall informative. In their open-ended comments, members of all stakeholder groups described concerns that the bulleted critique format leaves open the potential for transmitting incomplete information and/or insufficient detail. Applicants' responses to questions about summary statements were strongly influenced by review outcomes. Applicants, whose applications were not discussed, more often disagreed that summary statements helped them focus on problem areas that could be corrected and understand why the application was not discussed. Applicants, whose applications were discussed, agreed most often that summary statements helped them focus on problem areas and understand the Overall Impact score.

Most applicants, reviewers, POs and SROs rated the single resubmission policy, which limits applicants' opportunity to resubmit applications without fundamental revision, as having hindered the NIH peer review process. Respondents stated in their open-ended comments that some applicants, such as New Investigators and investigators with smaller research programs and/or limited research facilities, are disproportionately affected by the policy. NIH leaders continue to monitor the policy and its implications.

This report focuses on the aspects of the NIH peer review process identified by the surveys to be in need of further attention. However, responses to overall satisfaction and other summative questions indicate that most of the changes implemented during the Enhancing Peer Review initiative are rated positively by NIH stakeholders and are achieving their intended objectives. Most respondents rated the peer review system as fair and rate themselves as satisfied with the NIH peer review process.

Table of contents

Introduction	3
Changes Introduced after the Phase I Surveys Were Deployed	4
Shortened Applications	4
Narrative Overall Impact Statement	6
Changes Indicated by Stakeholders to be in Need of Further Attention	9
Nine-point Scoring Scale	9
Bulleted Comments	12
Single Resubmission Policy	17
Overall Satisfaction	19
Overall Ratings of the Enhancing Peer Review Changes by NIH Staff Members	25
Discussion	27
Appendix 1. Summary of Enhancing Peer Review changes	29
Appendix 2. Details about Sampling and Analysis	31
Appendix 3. Acknowledgements	35

Report on the Results of the Enhancing Peer Review Surveys: Phase II

In 2007, the National Institutes of Health (NIH) embarked on a self-study¹ of its peer review process to examine the effectiveness of this process, in light of the rapidly evolving scientific and public health landscape. The result of the NIH peer review self-study was a series of modifications to the NIH grant application and the peer review process designed to achieve three implementation goals: 1) Engage the Best Reviewers; 2) Improve the Quality and Transparency of Review; and 3) Ensure Balanced and Fair Reviews. The Enhancing Peer Review changes implemented over two years are summarized in [Appendix 1](#).

The fourth implementation goal, also recommended by the NIH peer review self-study, was the continuous review of peer review, a dynamic assessment of the outcomes brought about by the Enhancing Peer Review changes. In addition to periodic opinion surveys of key peer review stakeholders, continuous review of peer review also includes quantitative analyses of scoring patterns, applications submissions and award patterns.

The Enhancing Peer Review surveys were conducted early in fiscal years (FYs) 2010 and 2012 to assess the overall state of the NIH peer review system and the effect of the changes on five groups of stakeholders whose work is directly impacted by peer review: applicants, peer reviewers, Scientific Review Officers (SROs), Program Officials (POs) and members of the advisory councils and boards of the 24 NIH institutes and centers (ICs) that make research grant awards. This report summarizes the key results of the second (Phase II; conducted in FY 2012) surveys by presenting highlights from the Phase II surveys, as well as comparative analyses of responses to questions included on both the Phase I (FY 2010) and Phase II surveys.

The survey questions were developed to examine whether the changes enacted as part of the Enhancing Peer Review initiative had affected the peer review process as intended. Additional questions were developed to assess whether the fundamental purposes of peer review had been affected. Focus groups were conducted with SROs and POs prior to drafting the surveys, and field testing was conducted after the surveys were finalized to ensure the survey questions were relevant and clearly worded. In addition to the structured questions on the survey, open-ended questions were also available for respondents to enter comments on any topic they chose. These comments were tallied into primary categories and used to inform the responses to structured survey questions.

The surveys were programmed as web-based instruments and hosted on secure, confidential web sites by Research Triangle Institutes, International (RTI). RTI also prepared randomized, stratified samples of applicants and reviewers, distributed invitations to potential respondents, and collected and analyzed the survey data.

Number of Respondents: 265 SROs and 378 POs responded to the surveys. 882 applicants, 836 reviewers, and 175 Advisory Council members responded to the surveys. The sampling strategies and survey response rates for all stakeholder groups are shown in [Appendix 2](#); calculated bias estimates for the applicant and reviewer respondents are also shown in Appendix 2.

Results of the Phase II surveys for each stakeholder group are discussed in the following sections:

- Changes Introduced after the Phase I Surveys Were Deployed
- Changes Indicated by Stakeholders to be in Need of Further Attention
- Overall Satisfaction

¹ National Institutes of Health 2007-2008 Peer Review Self-Study:
<http://enhancing-peer-review.nih.gov/meetings/NIHPeerReviewReportFINALDRAFT.pdf>

Changes Introduced after the Phase I Surveys Were Deployed: Shortened Applications

The [page limits for NIH grant applications](#) were shortened beginning with all applications submitted for FY 2011 funding (submission deadlines in January 2010). The format of the shortened application was aligned with the NIH review criteria. The standard research strategy for an investigator-initiated research project grant was set at 12 pages, and in applications for shorter activities, such as NIH Small Grants (R03) or Exploratory/Developmental Grants (R21) the research strategy section was set at a 6-page limit. Other application sections, including the biographical sketch and the introduction to resubmission were also shortened. New [limits on Appendix materials](#) were established in 2007, and guide notices were issued in [2010](#) and [2011](#) reminding grantees that the appendix limits were unchanged under the shorter application format. The purpose of shortening the application was to shift the focus of the application to scientific impact and uniqueness/originality, placing reduced emphasis on standard methodological details. This change was also anticipated to reduce burden so that reviewers could read more of the applications assigned to the meeting.

Applicants and Reviewers

More applicants agreed than disagreed that the shortened application sections were sufficient for describing their research project ([Figure 1](#)). However, fewer applicants agreed that the Introduction to Resubmission was sufficient.

More reviewers agreed than disagreed that the shortened application sections were sufficient for evaluating the scientific merit of most research grant applications ([Figure 1](#)). However, fewer reviewers agreed that the Appendices section was sufficient.

SROs and POs

More SROs and POs agreed than disagreed that the shortened application sections were sufficient for evaluating scientific merit ([Figure 2](#)). However, fewer SROs and POs agreed that the Appendices were sufficient, and fewer POs agreed that the Research Strategy section was sufficient for clinical research applications.

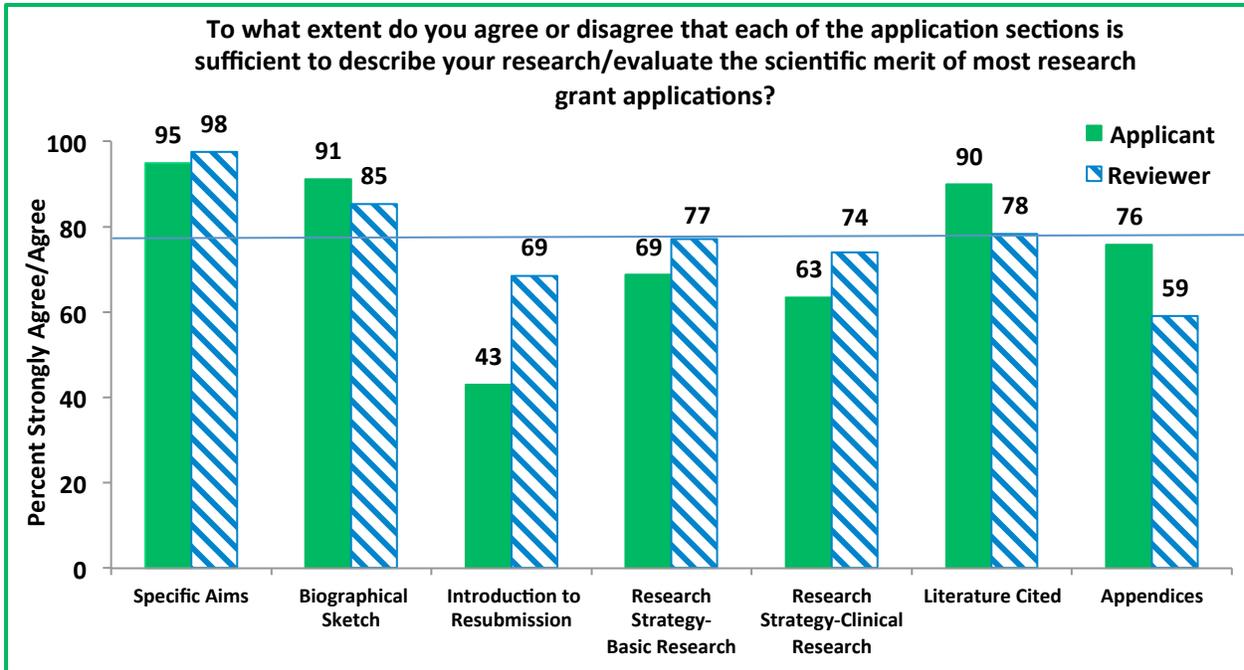


Figure 1. Applicant and reviewer ratings of whether each of the shortened applications sections is sufficient to describe the proposed research (applicants) or evaluate the scientific merit (reviewers) of research grant applications. The horizontal line on the graph indicates the grand mean level of agreement across all shortened applications sections for both applicants and reviewers (77%).

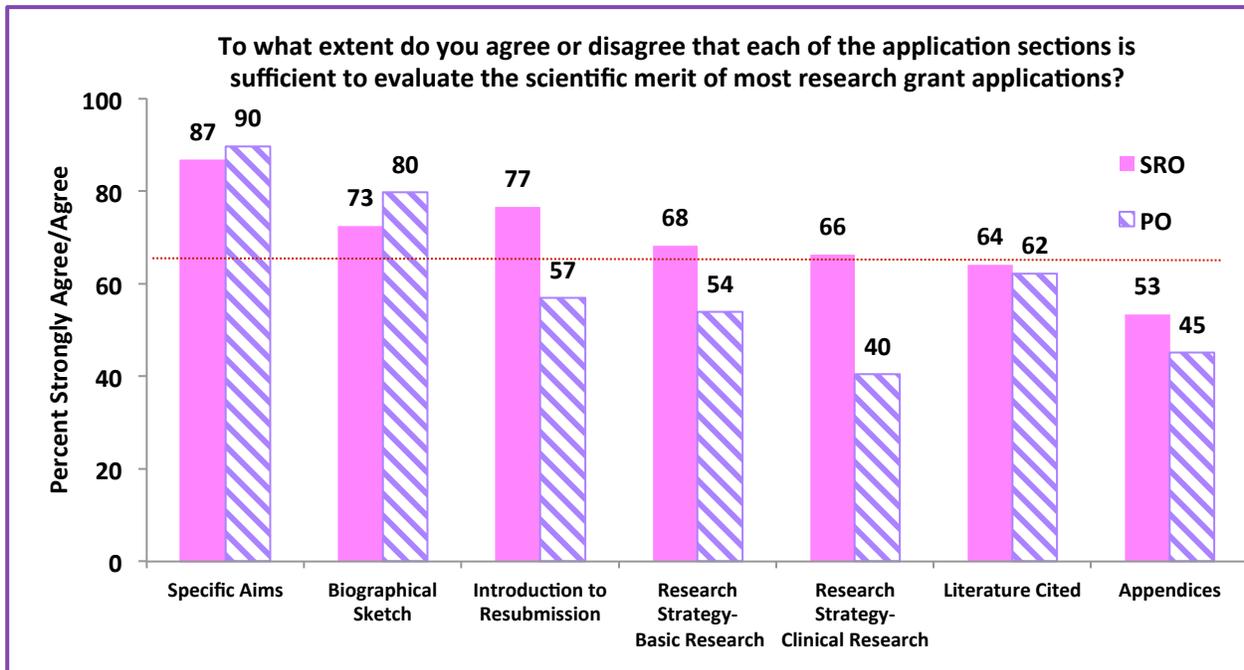


Figure 2. SRO and PO ratings of whether each of the shortened application sections is sufficient to evaluate the scientific merit of research grant applications. The horizontal line on the graph indicates the grand mean level of agreement across all shortened applications sections for both SROs and POs (65%).

Changes Introduced after the Phase I Surveys Were Deployed: Narrative Overall Impact Statement

In 2010, NIH responded to feedback from numerous sources, including the Phase I Enhancing Peer Review surveys, about the quality of information contained in summary statements. The format of the Overall Impact section in the critique template was modified from bulleted to narrative format and reviewers were instructed to write a paragraph summarizing the factors that informed their Overall Impact score.

Applicants

Applicants responded to a series of questions about the helpfulness of summary statements and other information provided by NIH for understanding their Overall Impact score.

- Overall, more applicants rated the following sections very helpful or somewhat helpful than rated them not very helpful or not at all helpful: Resume and Summary of Discussion (69%), reviewers' critiques (66%), Overall Impact statement (55%) and discussion with their Program Officer (56%).
- Applicants who reported that their applications had been funded were significantly more likely to rate all of the items very helpful or somewhat helpful for understanding their Overall Impact score in comparison to applicants whose applications were discussed but not funded ([Figure 3](#)).

Reviewers

- More reviewers agreed than disagreed that the narrative Overall Impact statement was helpful (78%) for communicating how the review criteria had contributed to the Overall Impact score.
- More reviewers also agreed than disagreed that the narrative Overall Impact statement was helpful for communicating why the application was not discussed (63%; [Figure 4](#)).

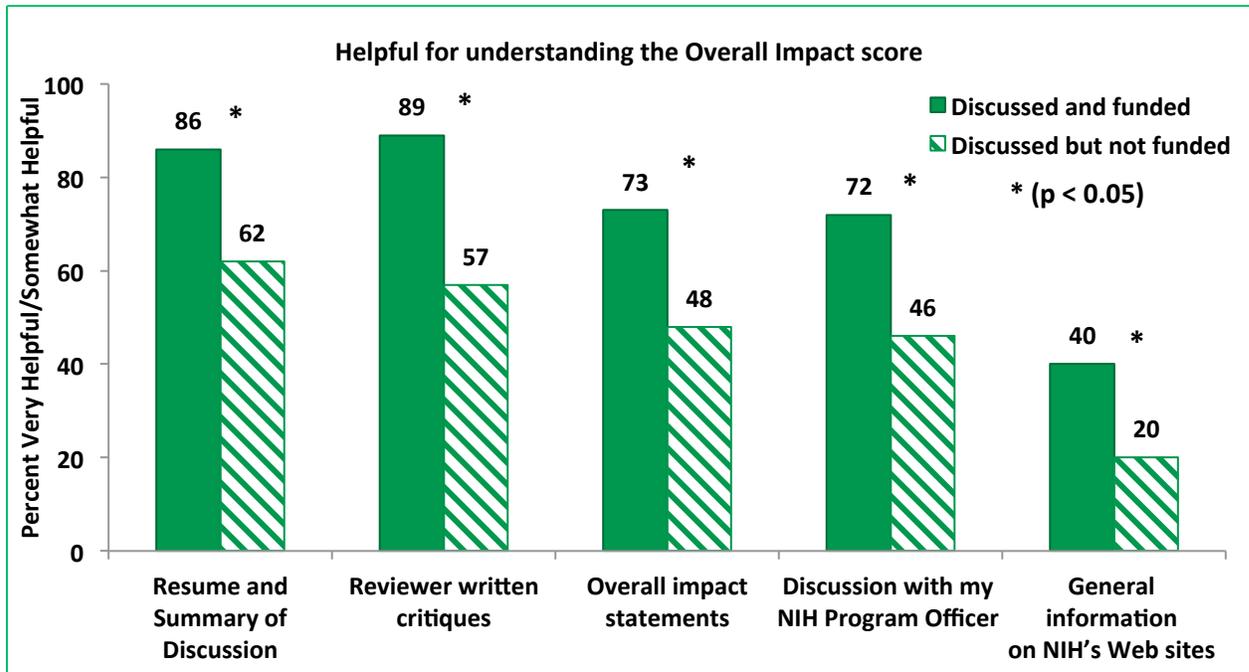


Figure 3. Applicant responses to questions about the helpfulness of information from Summary statements and other NIH sources for understanding their Overall Impact score. Applicants rated the information very helpful or somewhat helpful significantly more often if their application was funded than if their application was discussed but not funded.

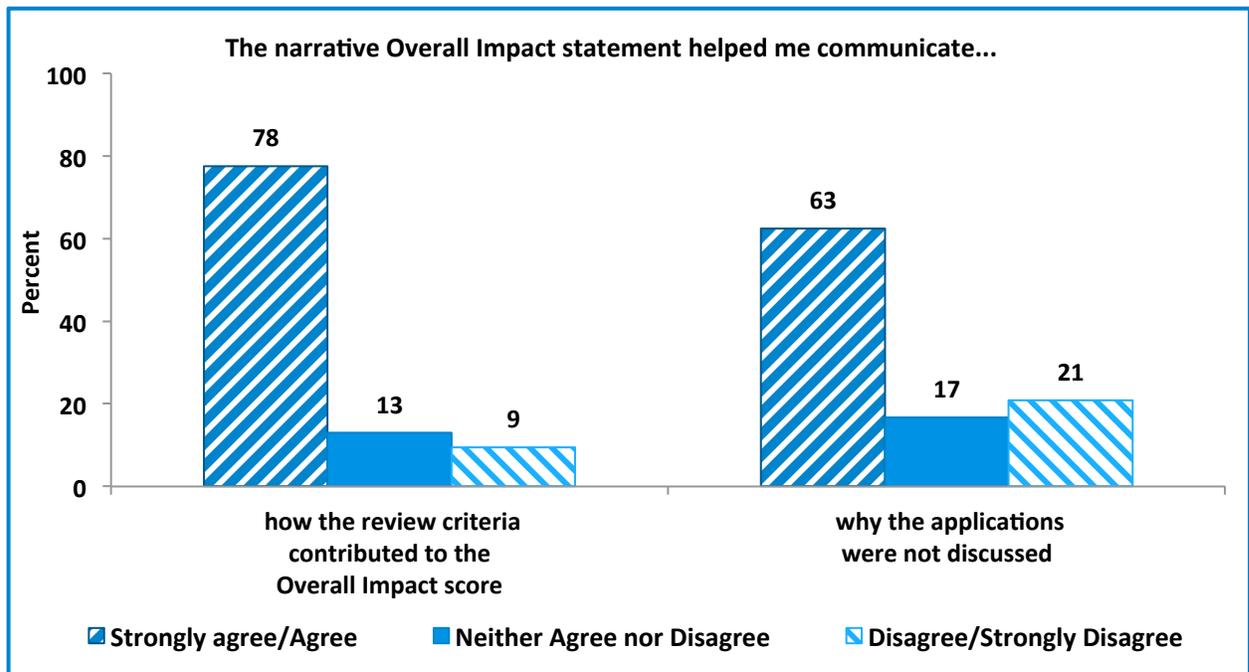


Figure 4. Reviewer responses to questions about whether the narrative Overall Impact statement was helpful for communicating how the strengths and weaknesses in each of the scored review criteria had contributed to the Overall Impact Score, and why applications were not discussed.

SROs and POs

Similar to reviewers, 67% of SROs agreed that since the introduction of the narrative Overall Impact statement, critiques had generally been helpful for understanding how the review criteria contributed to the Overall Impact score (Figure 5). Fewer SROs agreed (57%) that the narrative Overall Impact statement was helpful for understanding why applications were not discussed.

Fewer POs (41% and 45%, respectively; Figure 5) agreed with these statements. In a separate question, POs were asked to select up to three elements of the Enhancing Peer Review initiative they found most helpful for advising applicants after review. The most common selection, chosen by 42% of POs, was the narrative Overall Impact statement.

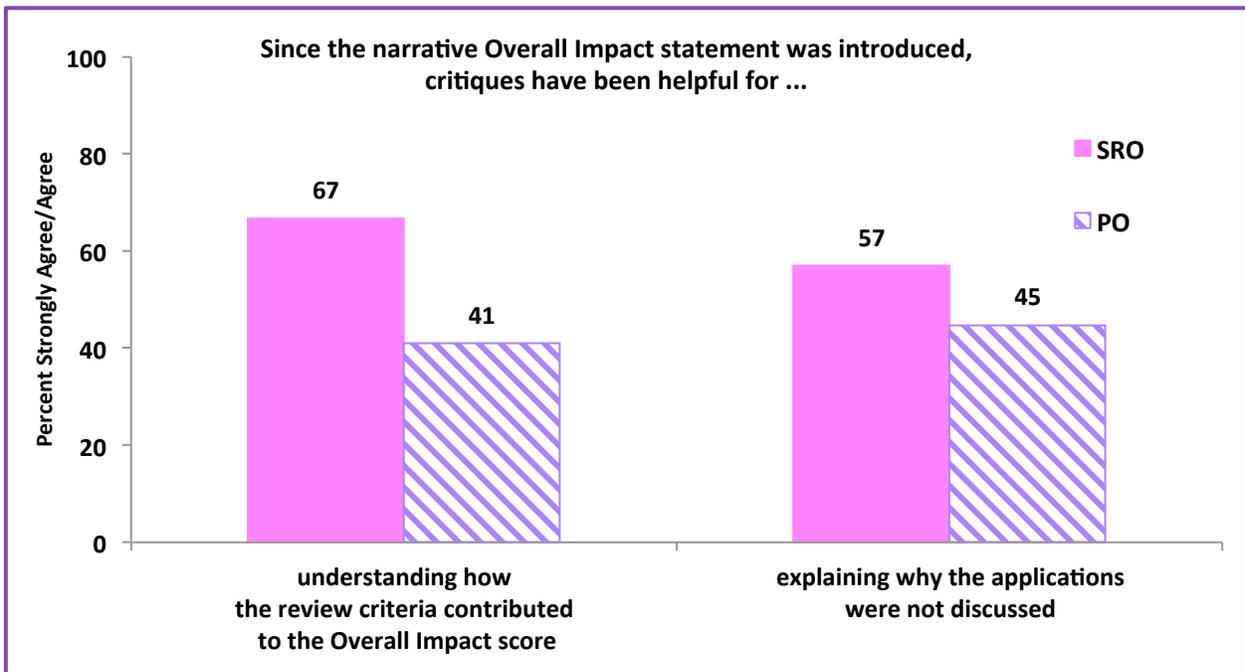


Figure 5. SRO and PO responses to questions about whether the narrative Overall Impact statement was helpful for understanding how the scored review criteria had contributed to the Overall Impact score, and why applications were not discussed.

Changes Indicated by Stakeholders to be in Need of Further Attention: Nine-Point Scoring Scale

Reviewers

- Most reviewers reported in Phase II that the nine-point scoring scale had sufficient range to communicate meaningful differences in the quality of applications. This response was similar to the one given in the Phase I surveys ([Figure 6](#)).
- Of the 548 reviewers who entered free-form comments at the end of their surveys, 97 (18%) commented on the nine-point scoring system. Most of these comments were centered on concerns about score compression or inflation, or their observation that a disproportionate number of scores were given in a specific area of the score distribution. Many reviewers expressed a need for stronger coaching from review staff and/or for more training to encourage reviewers to spread their scores. Other reviewers' comments expressed a need for half-point increments to be added to the score distribution.

SROs and POs

- SROs agreed significantly less often in Phase II than in Phase I that the assigned reviewers use the full range of scores in their preliminary score assignments ([Figure 7](#)).
- Program Officers were asked to select up to three elements of the Enhancing Peer Review initiative they found LEAST helpful for advising applicants after review. POs most often chose the single resubmission policy (59%), bulleted comments (49%) and the nine-point scoring scale as least helpful for advising applicants (36%) ([Figure 8](#)).

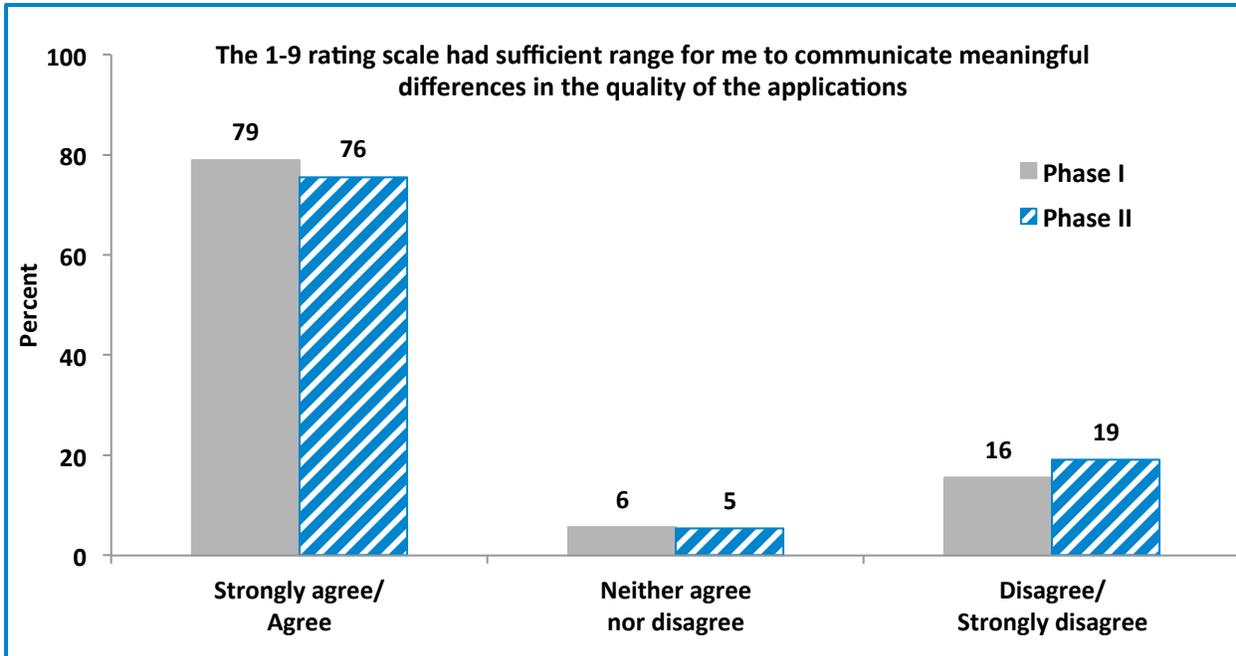


Figure 6. Reviewer responses to a question about whether the nine-point scale had sufficient range to communicate meaningful differences in the quality of applications. There were no significant differences between reviewer responses on the Phase I and Phase II surveys.

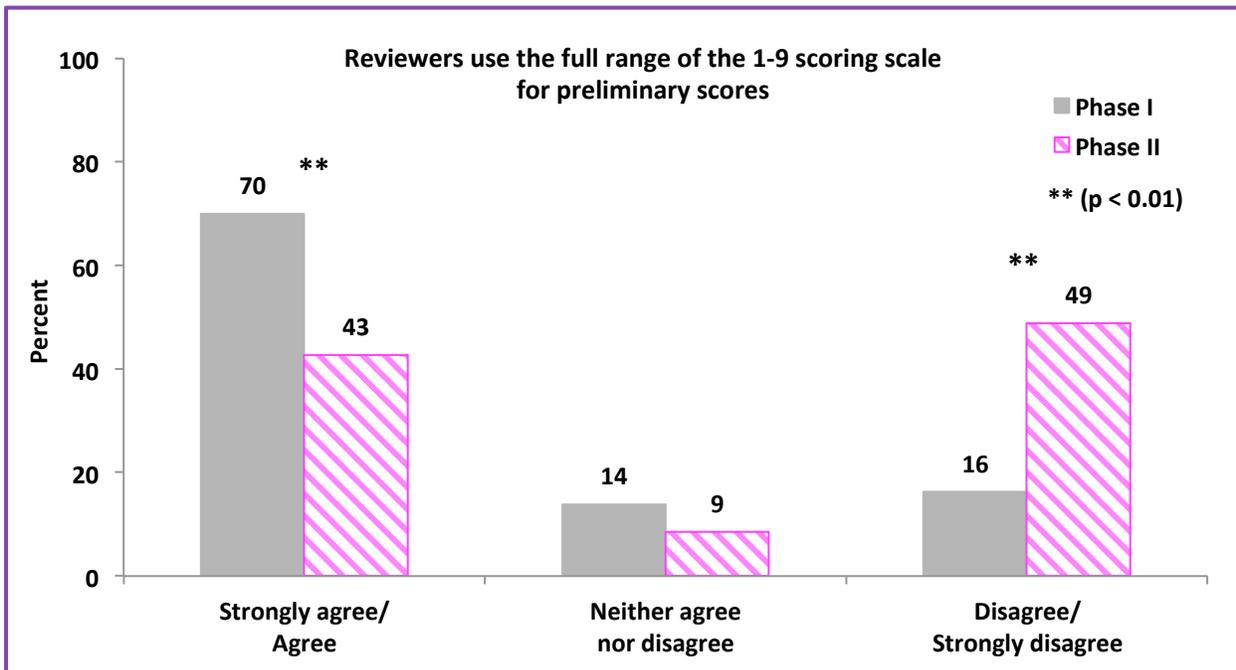


Figure 7. SRO responses to a question about whether the reviewers are using the full range of scores in their preliminary scoring of their assigned applications. Significantly fewer SROs agreed that reviewers use the full range of scores in Phase II in comparison to Phase I.

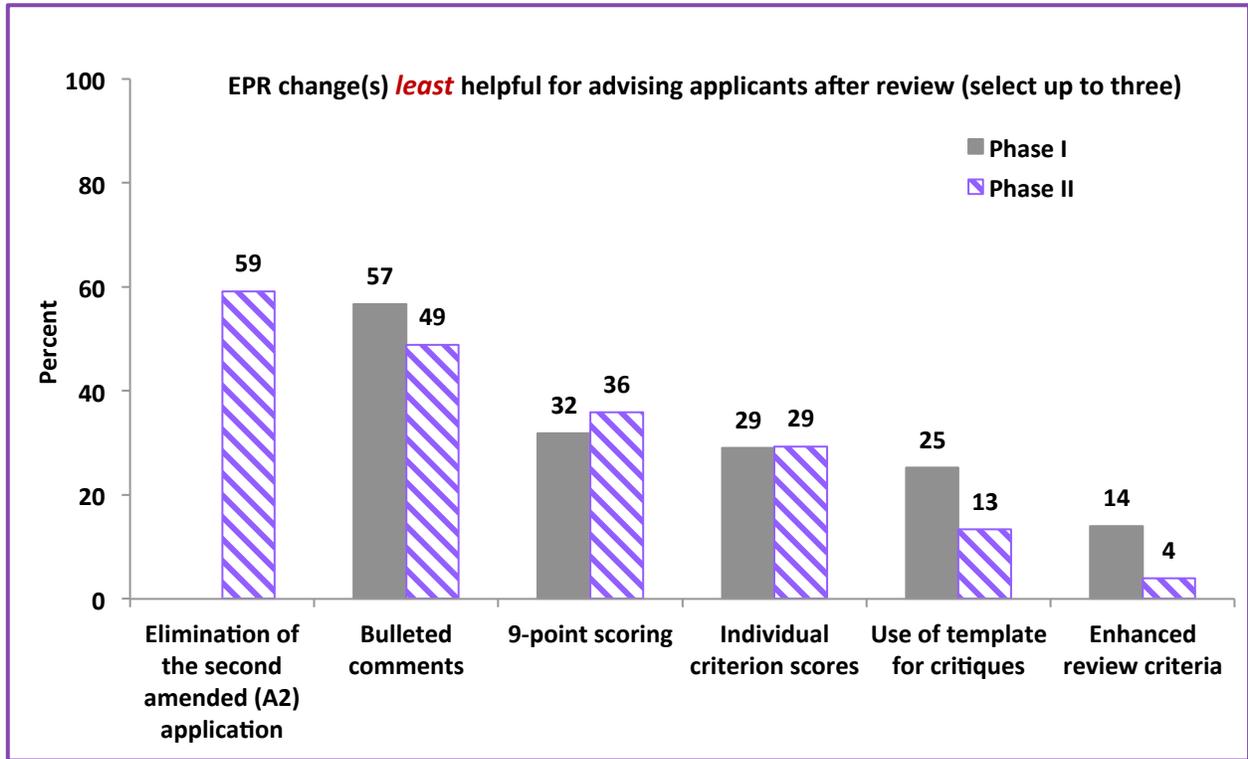


Figure 8. POs were asked to select up to three of the Enhancing Peer Review changes they found *least* helpful for advising applicants after review. The percent of POs who selected each of the six most commonly chosen EPR changes in Phase I and Phase II is shown. There was a modest improvement in the rates of selection in Phase II for most changes in comparison to Phase I; however, the proportion of POs who selected 9-point scoring as least helpful increased slightly in Phase II. Elimination of the second amended application, a change that was made effective after the Phase I surveys, was selected most often by POs in Phase II.

Changes Indicated by Stakeholders to be in Need of Further Attention: Bulleted Comments

The format of the comments on each scored criterion within the critique templates was modified in the Enhancing Peer Review initiative to a bulleted list of strengths and weaknesses. The purpose of the change to a bulleted format was to discourage reviewers from summarizing the application and/or making suggestions about potential methodological improvements; the purpose of the bulleted critique format is to encourage concise statements about factors that affect scientific merit.

Applicants

The applicant survey contained a number of general questions about the helpfulness of summary statements for understanding the outcome of the review of their application.

- As indicated in [Figure 3](#), applicants rated reviewer critiques very helpful or somewhat helpful more often than they rated them not helpful for understanding their Overall Impact score; applicants whose applications were funded rated the critiques as helpful significantly more often than did applicants whose applications were discussed but not funded.
- Applicants disagreed (49%) more often than they agreed (38%) that summary statements were helpful for understanding why their application was not discussed ([Figure 9](#)).
- Most applicants whose applications were not funded (both discussed and not discussed) and were not resubmissions agreed that summary statements helped them decide whether to resubmit their application (67%; [Figure 10](#)).
- Applicants whose applications were funded agreed that the summary statement helped them focus on problem areas that could be corrected significantly more often than those whose applications were not funded, or not discussed ([Figure 11](#)).

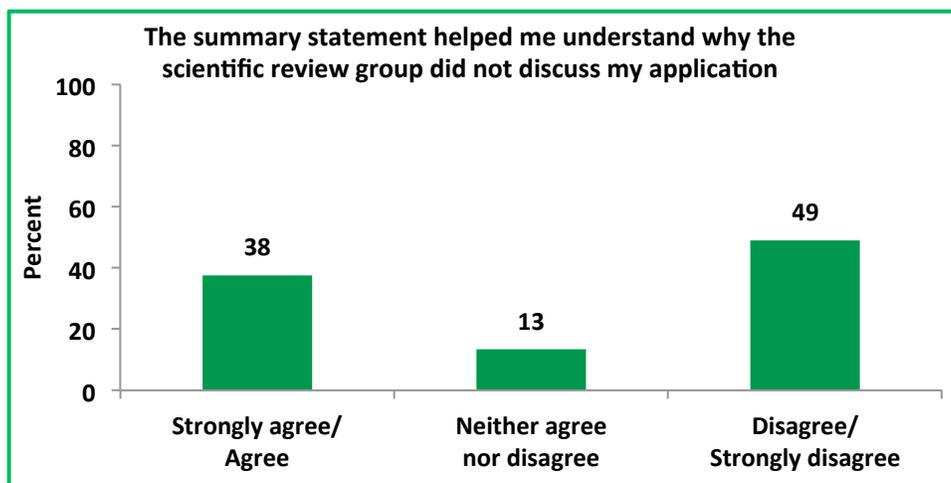


Figure 9. Applicant responses to a question about whether their summary statement was helpful for understanding why their application was not discussed. This question was offered to only applicants who had reported in a branching question that their most recent application had been not discussed.

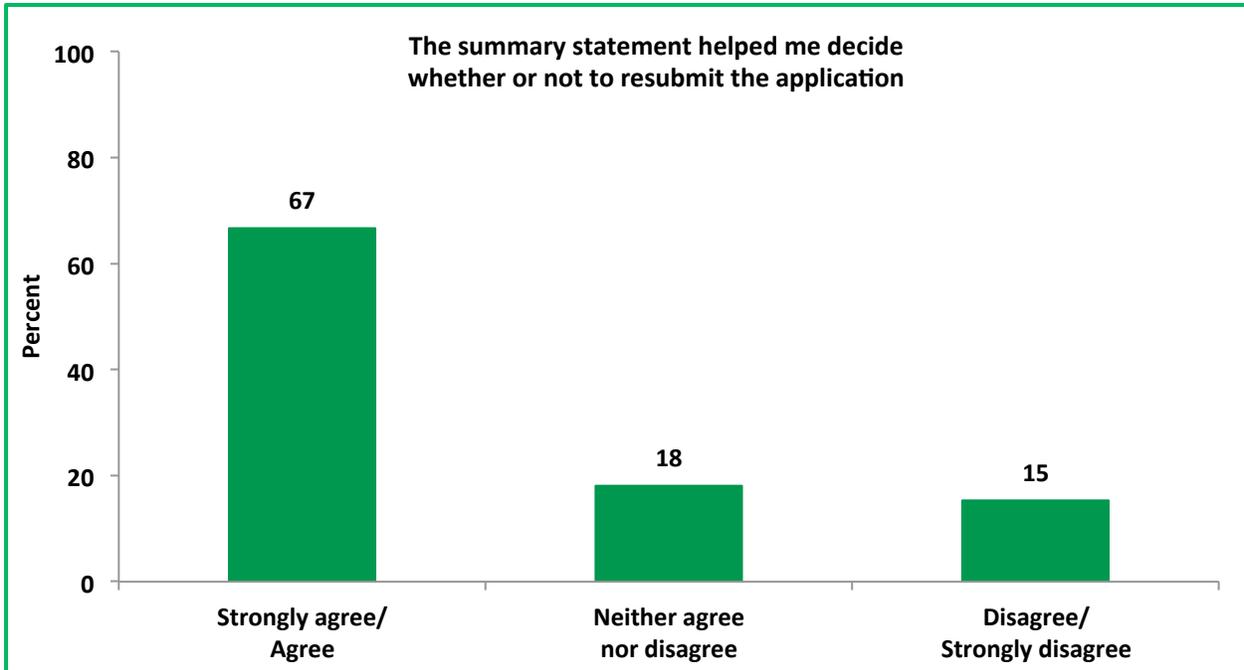


Figure 10. Applicant responses to a question about whether their summary statement helped them decide whether to resubmit their application. This question was offered only to respondents who reported that their most recent application was not funded and was not a resubmission.

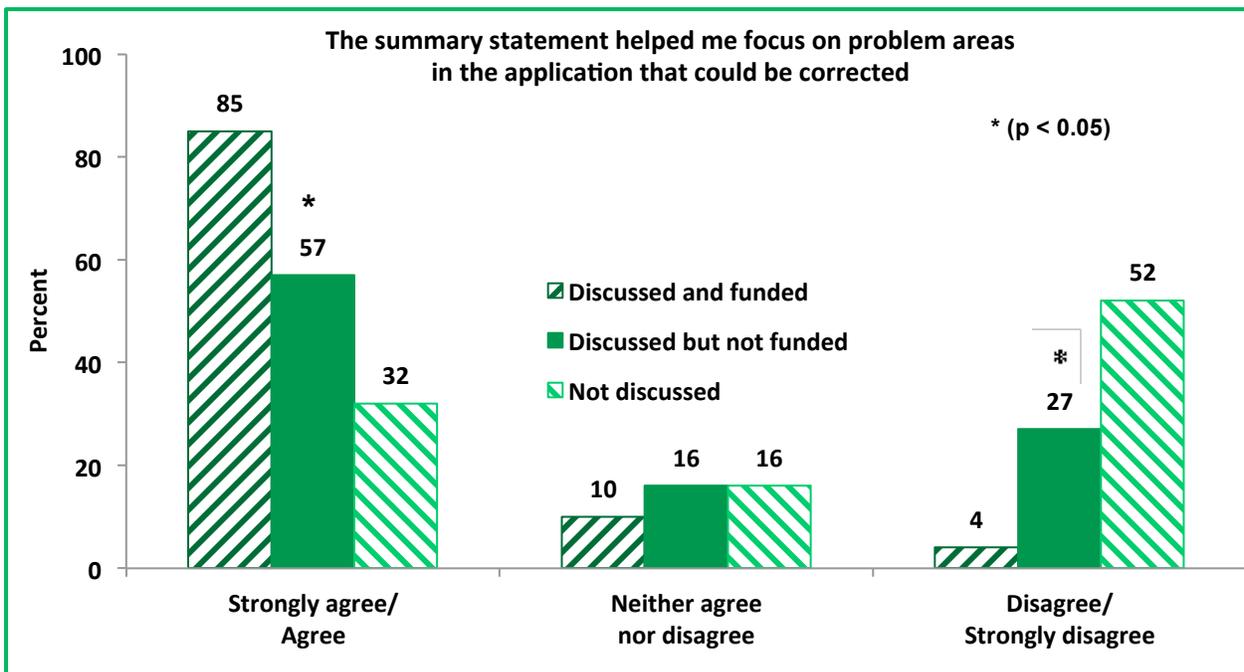


Figure 11. Applicants whose applications were funded agreed significantly more often than applicants whose applications were not funded that the summary statement helped them focus on correctable problems in the application.

*the three groups of applicants are significantly different from each other in a pairwise fashion.

SROs and POs

- SROs were asked a number of questions related to the reviewers' willingness to prepare critiques that were complete and conformed to the instructions they were provided. In most cases, SROs' responses to these questions in Phase II were significantly improved over Phase I. Even with this improvement, only 37% of SROs reported that the bulleted critiques were expressed in complete passages and 47% of SROs agreed that the bulleted comments generally cover all of the points made by reviewers during the discussions at the meeting ([Figure 12](#)).
- In Phase II, POs' ratings of the information contained in summary statements were significantly improved over Phase I. However, POs also reported low levels of agreement (24%) when asked whether the bulleted comments were expressed in complete thoughts ([Figure 13](#)). 58% of POs agreed that the critique content reflected the discussions at the meeting.

Advisory Council Members

- Advisory Council members who reported using summary statements as part of their council deliberations were also asked whether they agree that bulleted comments were expressed in complete, well-composed thoughts. They were nearly equally likely to agree (43%) as to disagree (37%) with this statement ([Figure 14](#)).
- Advisory Council members were also asked to rate the extent to which they found the bulleted comments helpful for understanding the scientific merit of the application under each of the five scored review criteria ([Figure 15](#)). Most Advisory Council members agreed that the bulleted comments were helpful for understanding the merit under Innovation (70%) and Investigator (70%), but fewer Advisory Council members agreed that the comments were helpful for understanding the merit under Approach (56%) and Significance (53%).

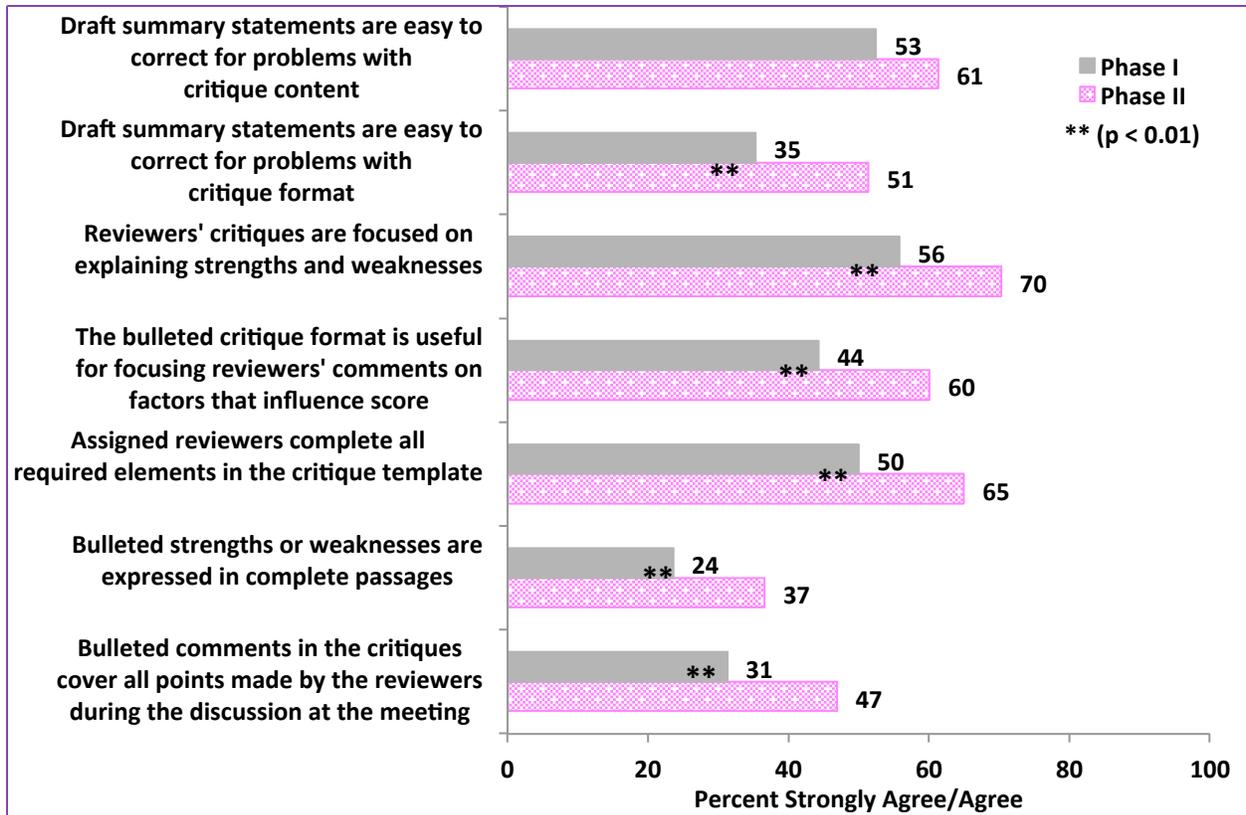


Figure 12. SRO responses to a series of questions about how well reviewers meet expectations in preparing their critiques. There was significant improvement on most of these benchmarks in Phase II in comparison to Phase I. However, most SROs did not agree that the bulleted comments are expressed in complete passages or that they cover all points made by reviewers during the discussions at the scientific review group meeting.

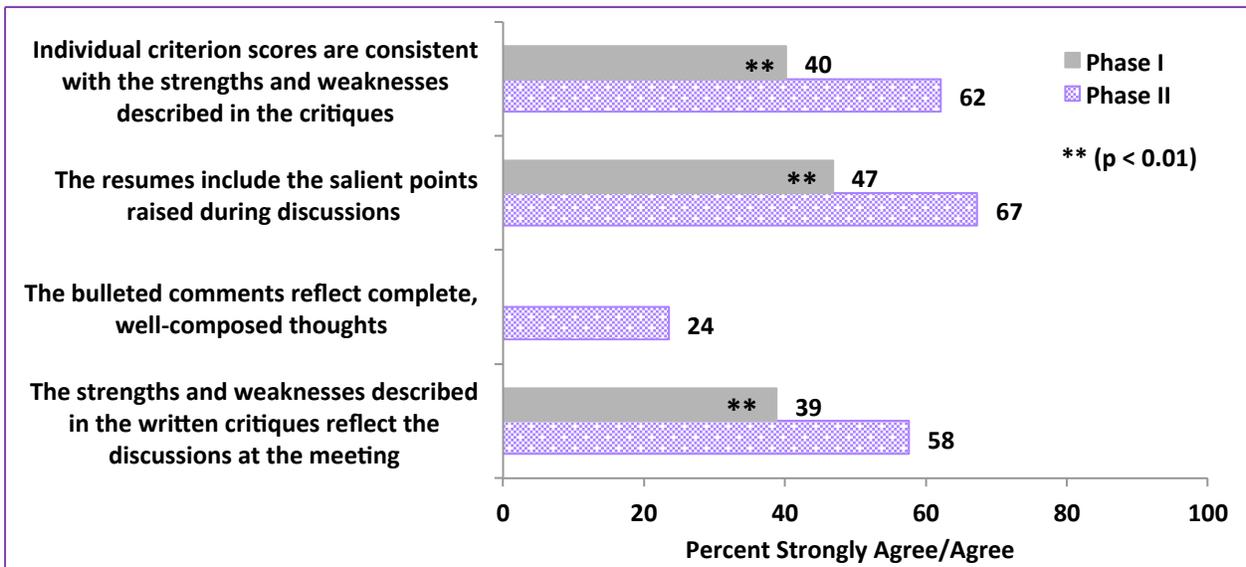


Figure 13. PO ratings of summary statements for applications assigned to their portfolio. There was significant improvement on most of the ratings in Phase II in comparison to Phase I. However, most POs did not agree that the bulleted comments are expressed in complete, well-composed thoughts (this question was only offered to POs on the Phase II survey).

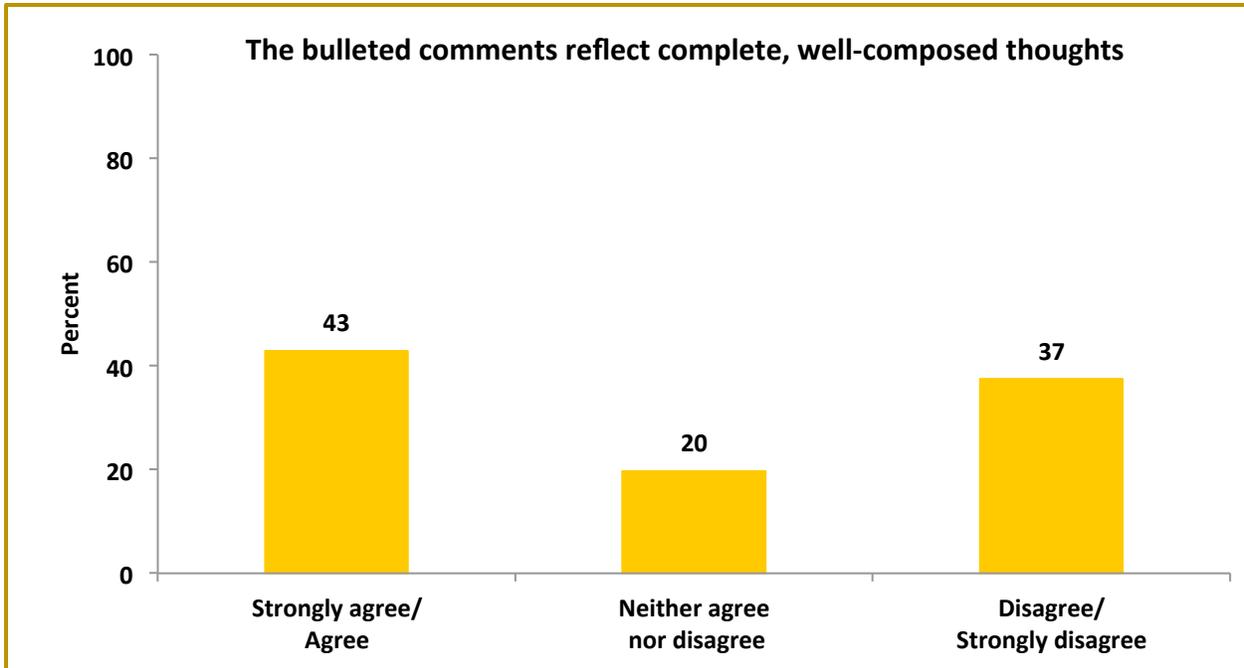


Figure 14. Advisory council member responses were equivocal when asked whether the bulleted comments reflected complete well-composed thoughts. Advisory council members were asked in a branching question about the types of materials they used as part of their advisory council deliberations. Only advisory council members who indicated that they used summary statements as part of their council deliberations were asked about the bulleted comments (n = 119).

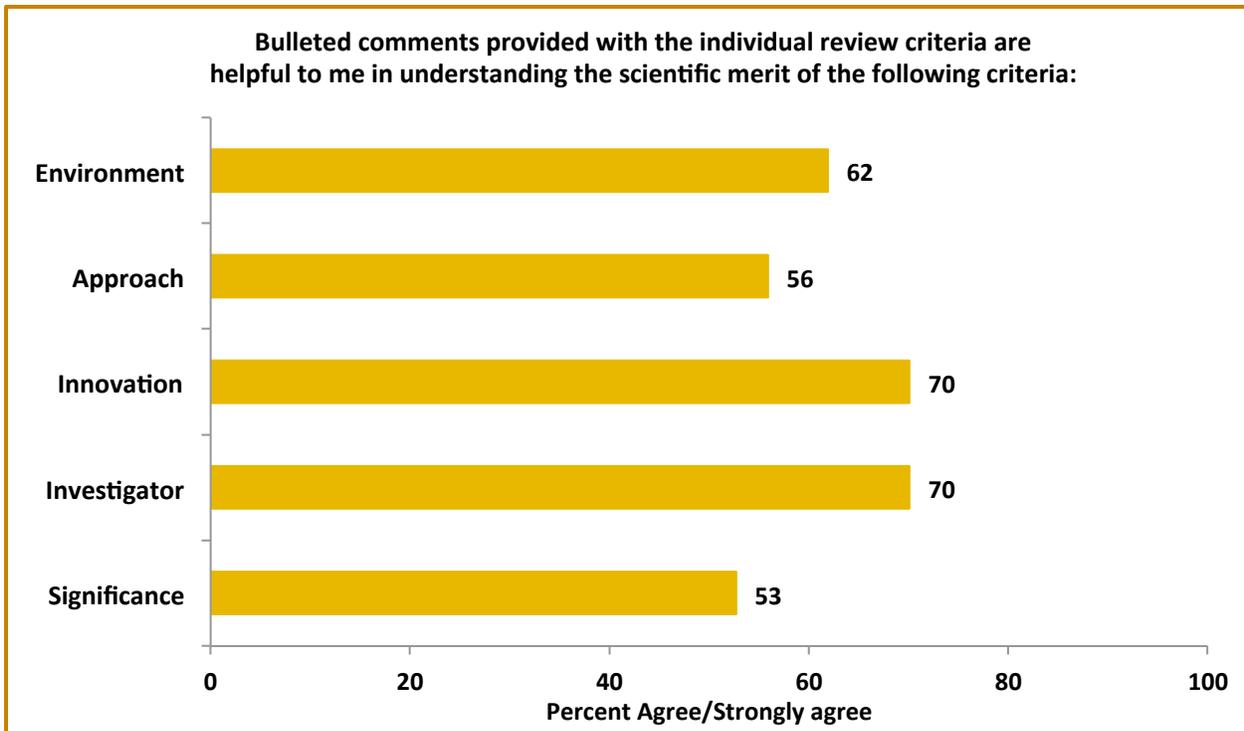


Figure 15. Advisory Council member ratings of whether the bulleted comments were helpful for understanding the scientific merit of each of the five scored review criteria.

Changes Indicated by Stakeholders to be in Need of Further Attention: Single Resubmission Policy

The NIH Peer Review self-study conducted in 2007 exposed a prevalent concern among grantees that the peer-review system had taken on the role of identifying potential improvements and instructing applicants of ways to improve the application. As a result, grantees perceived that a system of “queuing” had developed, where meritorious applications with relatively minor weaknesses were scored in a manner consistent with their potential for further improvement. A clear consensus expressed by applicants was that it was not the role of peer review to suggest methodological improvements; the peer review self-study reported the recommendations that 1) amended applications be eliminated, and 2) reviewers’ critiques be focused solely on scientific merit.

NIH elected to retain a single resubmission, rather than eliminate amended applications altogether. The [NIH Policy on Resubmission Applications](#) states that, beginning with applications submitted for the January 25, 2009 due dates (FY 2010 councils) and beyond, the NIH accepts only a single amendment to the original application. The purpose of this policy is to increase the numbers of high-quality original and first amendments that can be funded earlier.

Applicants and Reviewers

Applicants and reviewers were asked whether the single resubmission policy had helped, had no effect, or hindered the NIH peer review process. Respondents who selected “helped” or “hindered” were asked to briefly describe how the policy had helped or hindered the peer review process.

Most applicants responded that the single resubmission policy had hindered the NIH peer review process. Applicants who indicated in the survey that their most recent application had been funded were significantly more likely to respond that the single resubmission policy helped the peer review process; nevertheless, 50% of funded applicants responded that peer review was hindered by the policy ([Figure 16](#)). In their open-ended responses, 44% of applicants said the policy reduced the likelihood of having an application funded. 47% of applicants indicated that they had to submit more new applications to compensate for the loss of the second resubmission. Applicants voiced concerns about the provision that an application had to be “fundamentally revised” to be submitted as a new application after being reviewed as a resubmission. Applicants also expressed concern that the policy disadvantaged New Investigators and investigators with smaller laboratories, since it is more difficult for these investigators to change the direction of their research.

Most applicants who rated the single resubmission policy as having helped the peer review process stated in their open-ended comments that the policy had achieved its desired effect, as evidenced by the increased number of new applications now being funded and the overall reduction in the average time to award from first submission.

Most reviewers (55%) also rated the single resubmission policy as having hindered the peer review process ([Figure 17](#)). In their open-ended comments, reviewers expressed concerns similar to those of applicants described above. In addition, 38% of reviewers and 22% of applicants expressed the opinion that the role formerly played by peer review in suggesting improvements to research plans was valid and desirable.

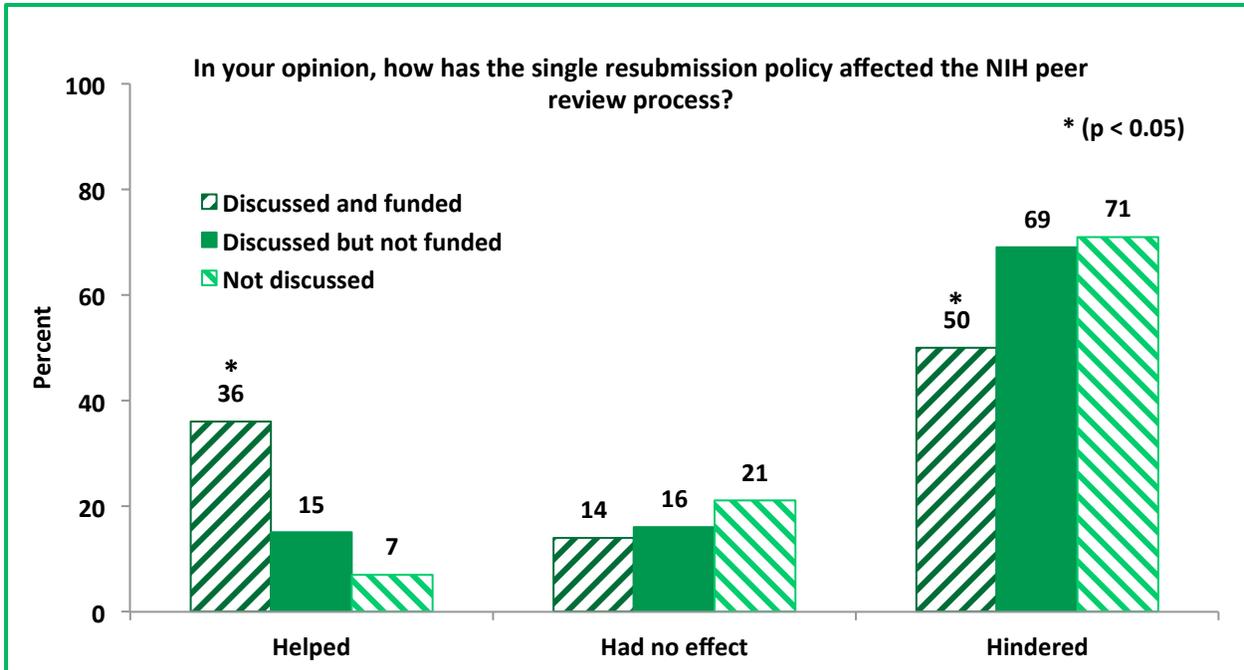


Figure 16. Most applicants rated the single resubmission policy as having hindered the NIH peer review process. Applicants whose applications were funded rated the policy as having helped the peer review process significantly more often and rated it as having hindered peer review significantly less often than applicants whose applications were not funded or not discussed.
* Applicants whose applications were discussed and funded are significantly different from the other two groups.

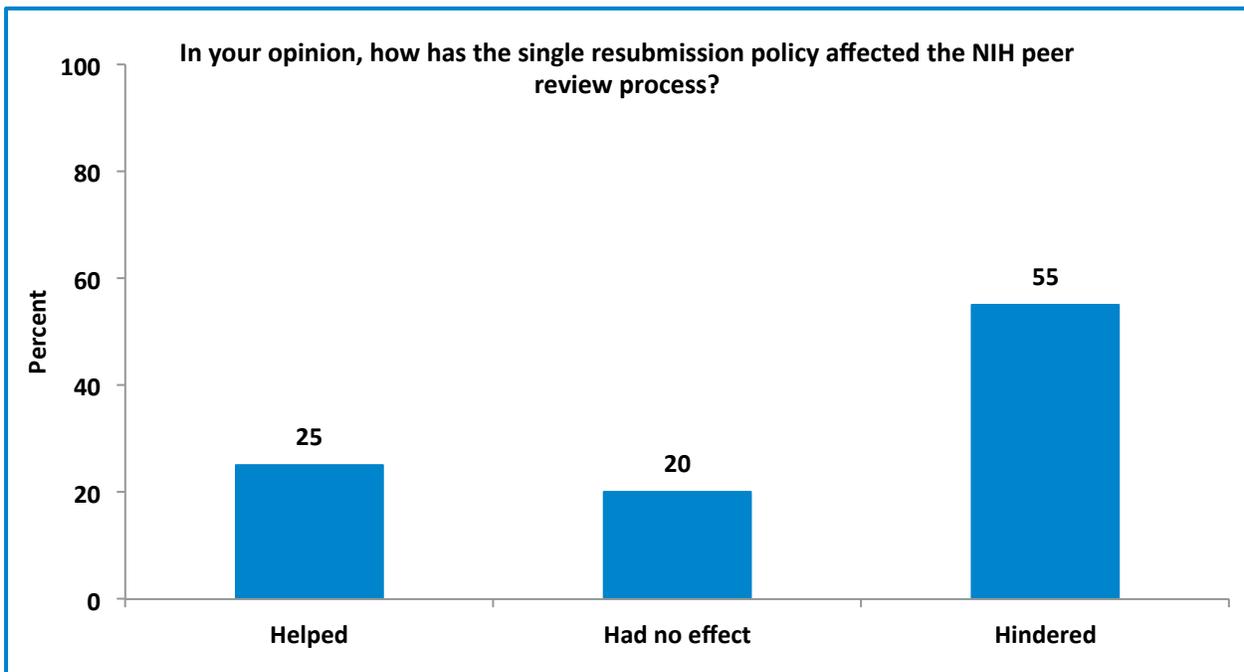


Figure 17. Most reviewers rated the single resubmission policy as having hindered the NIH peer review process.

Overall Satisfaction

All stakeholders were asked to respond to three overall satisfaction questions:

- 1) Overall, which peer review system do you prefer – the new system or the old system?
- 2) How fair is the peer review process at NIH?
- 3) How satisfied are you with the peer review process at NIH?

- Applicants reported a preference for the new peer review system over the old system significantly more often in Phase II than in Phase I ([Figure 18](#), top panel).
- Applicants continued to report the peer review system as fair or very fair more often than they rated it as unfair or very unfair. The improvement in their ratings in Phase II was not significant compared to Phase I ([Figure 18](#), center panel).
- Applicants reported they were satisfied or very satisfied with the peer review process significantly more often in Phase II than in Phase I ([Figure 18](#), bottom panel).
- All applicant satisfaction ratings were significantly affected by review outcome; applicants whose applications were funded were significantly more likely to rate the peer review system positively than those whose applications were not discussed.

- Most reviewers responded that they preferred the new peer review system over the old. Significantly more reviewers preferred the new peer review system in Phase II than in Phase I ([Figure 19](#), top panel).
- Most reviewers rated the peer review system as fair ([Figure 19](#), center panel) and rated themselves as satisfied ([Figure 19](#), bottom panel) in both phases and there were no significant differences between Phases in their ratings.

- Significantly more SROs reported a preference for the new system in Phase II than in Phase I ([Figure 20](#), top panel).
- Significantly more SROs rate the system as fair in Phase II than in Phase I ([Figure 20](#), center panel).
- Significantly more SROs reported themselves as satisfied with the peer review system in Phase II than in Phase I ([Figure 20](#), bottom panel).

- POs equally preferred the new versus the old peer review system in Phase II and their preference for the old system was significantly reduced from Phase I ([Figure 21](#), top panel).
- Most POs rated the peer review system as fair or very fair in both Phases ([Figure 21](#), center panel).
- Significantly more POs rated themselves as satisfied with the peer review system in Phase II than in Phase I ([Figure 21](#), bottom panel).

- Equal numbers of Advisory Council members preferred the old as did the new system in Phase II. They preferred the old system significantly more often in Phase II than in Phase I ([Figure 22](#), top panel).
- Most Advisory Council members (79%) rated the peer review system as fair ([Figure 22](#), center panel).
- Most Advisory Council members (72%) rated themselves as satisfied with the peer review system ([Figure 22](#), bottom panel).

Overall Satisfaction: Applicants

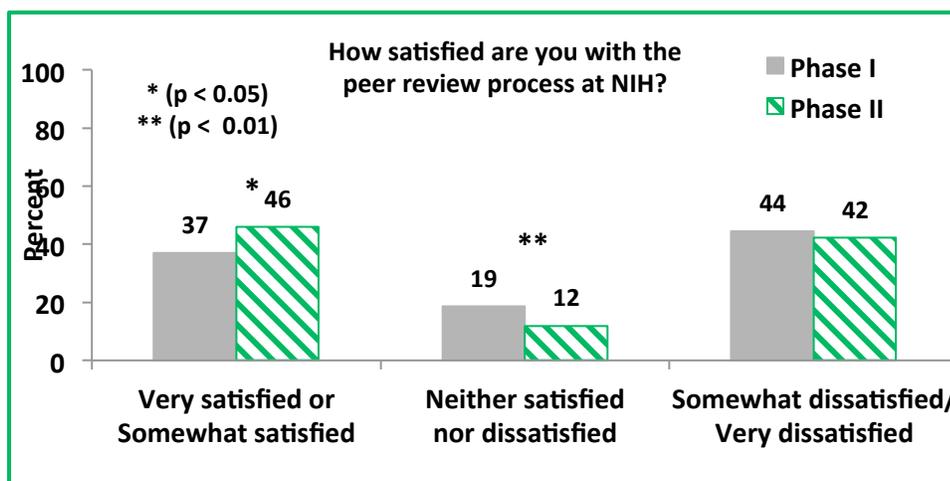
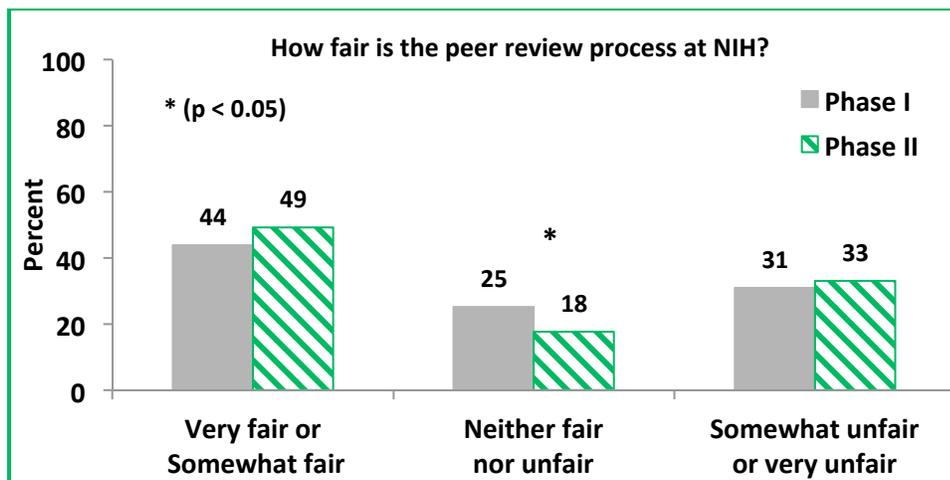
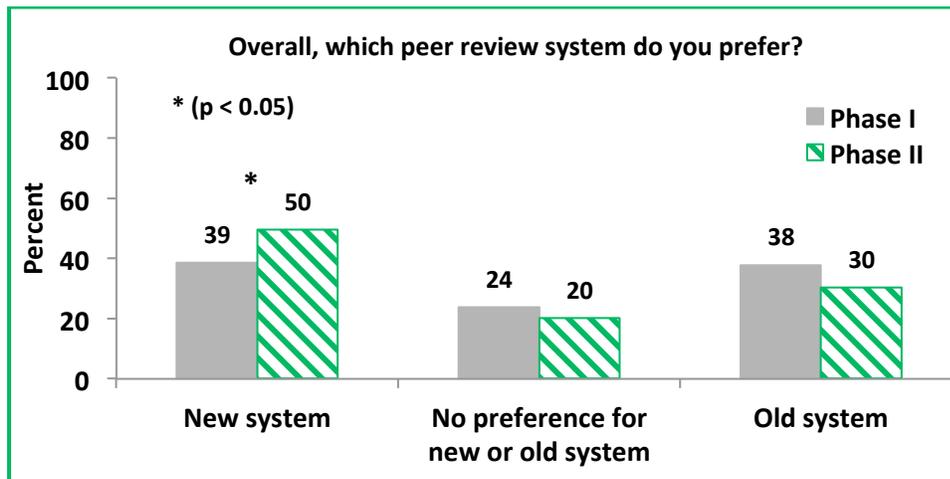


Figure 18. Applicant responses to the three overall evaluation questions on the Phase I and Phase II surveys. Significantly more applicants preferred the new peer review system and rated themselves as satisfied in Phase II than in Phase I.

Overall Satisfaction: Reviewers

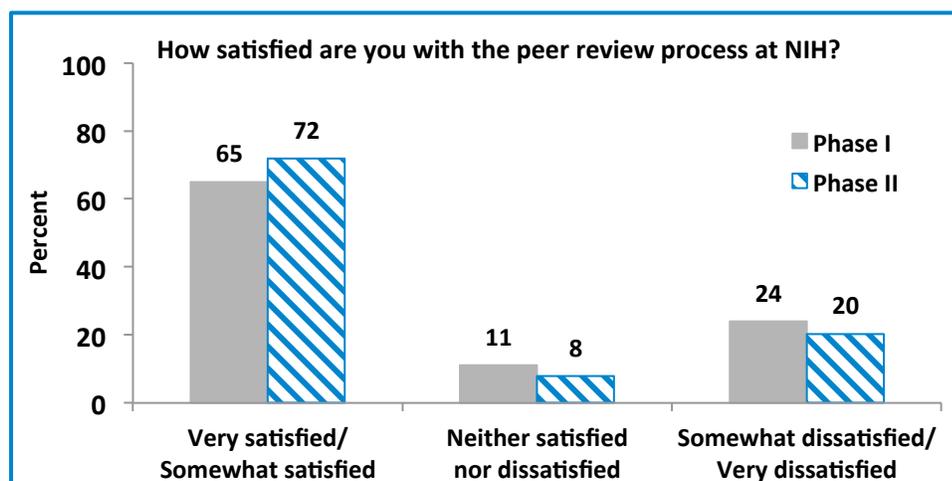
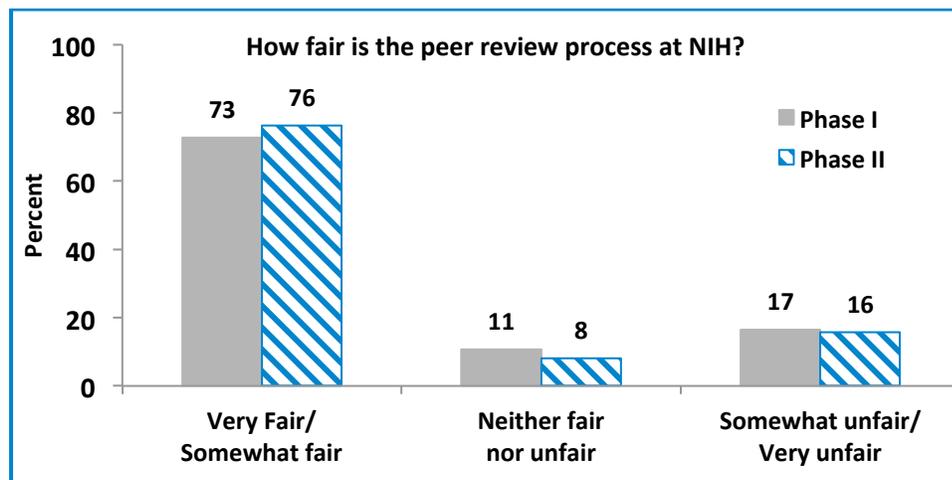
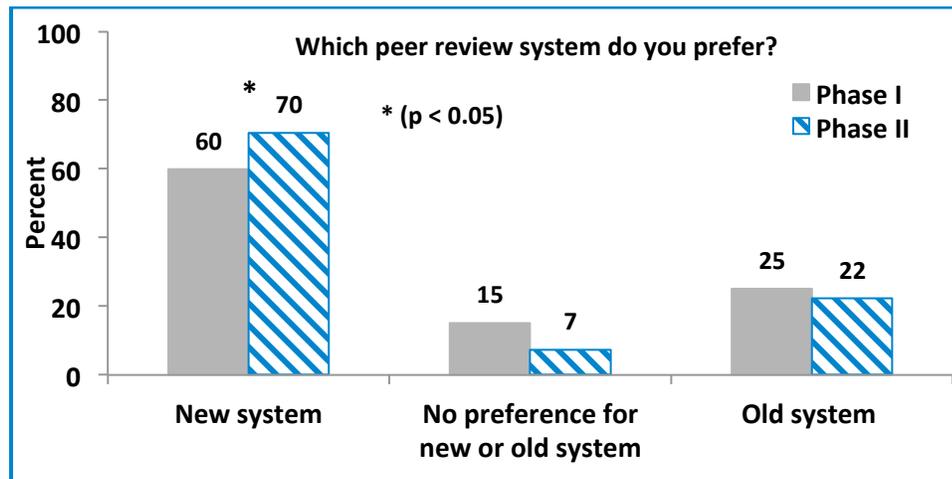


Figure 19. Reviewer responses to the three overall evaluation questions on the Phase I and Phase II surveys. Significantly more applicants preferred the new peer review system in Phase II than in Phase I.

Overall Satisfaction: SROs

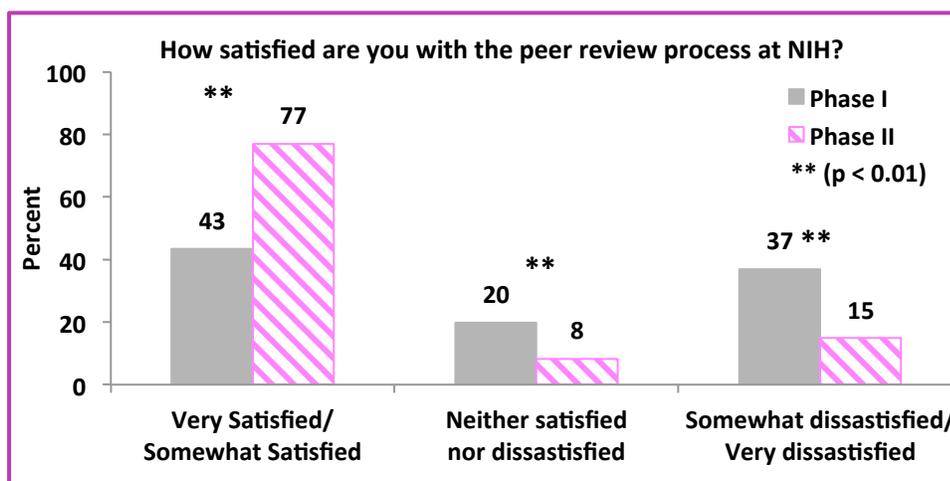
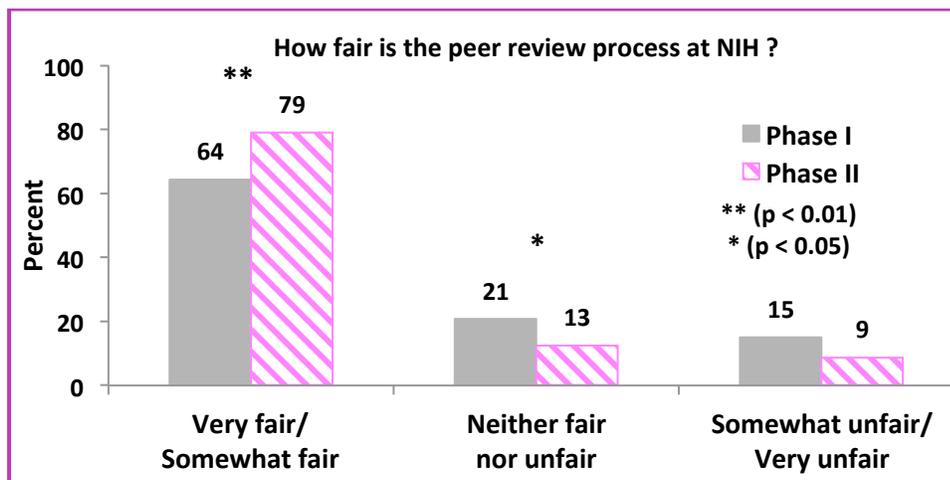
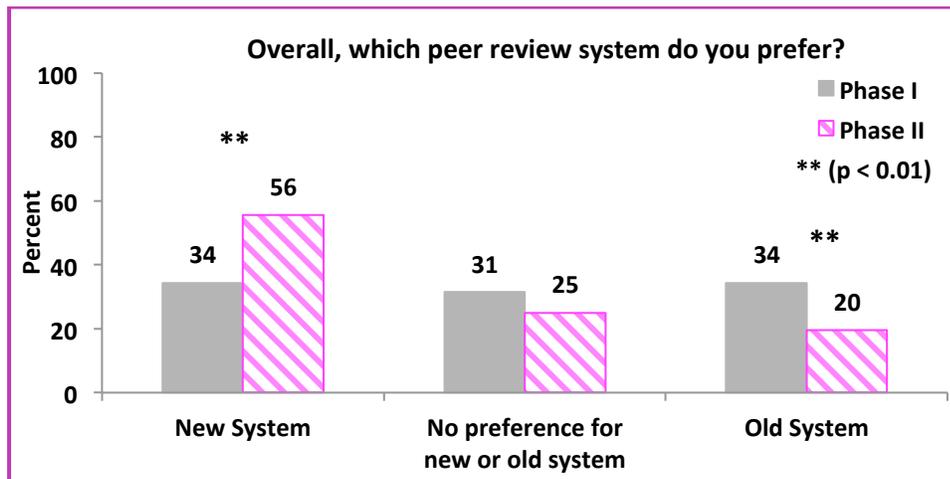


Figure 20. SRO responses to the three overall evaluation questions on the Phase I and Phase II surveys. Significantly more SROs preferred the new peer review system, rated the system as fair, and rated themselves as satisfied in Phase II than in Phase I.

Overall Satisfaction: POs

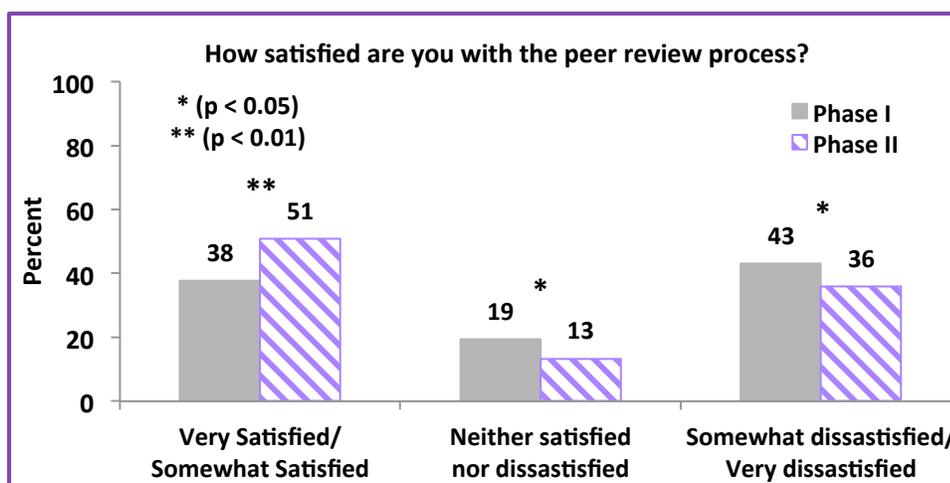
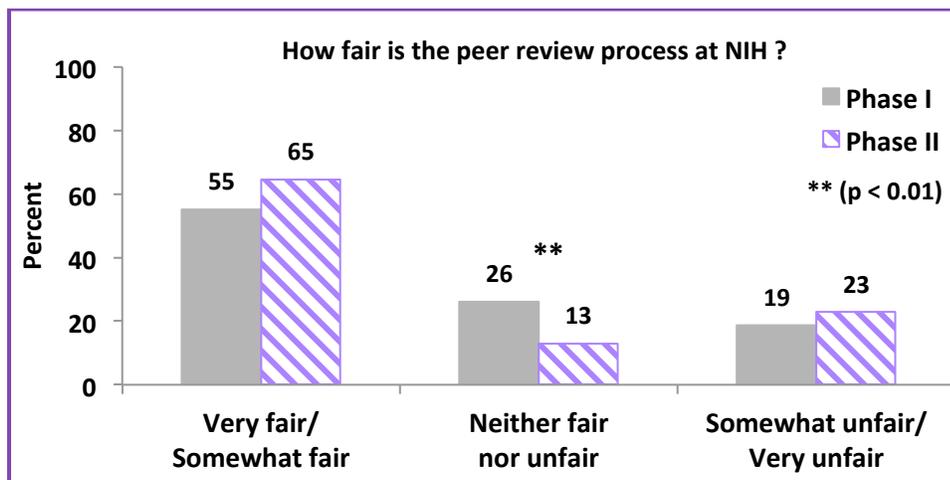
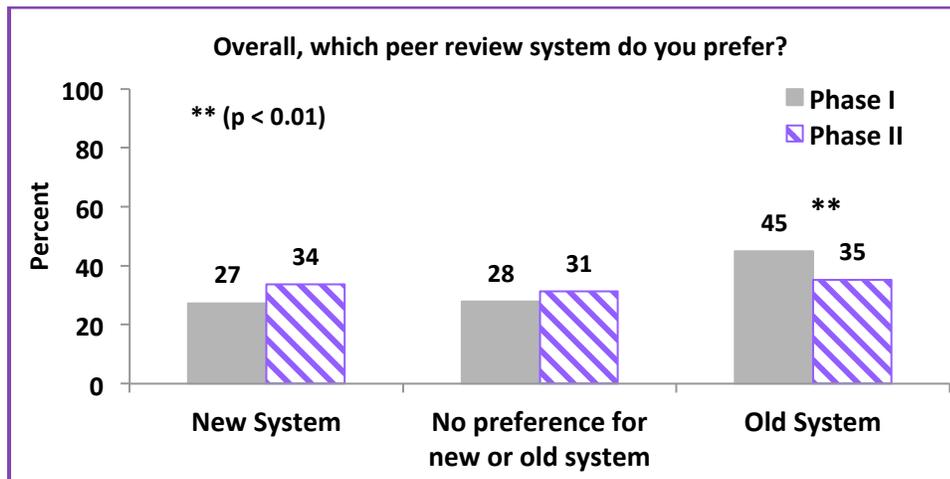


Figure 21. PO responses to the three overall evaluation questions on the Phase I and Phase II surveys. Significantly more POs rated themselves as satisfied in Phase II than in Phase I. However, other satisfaction ratings showed only modest improvement.

Overall Satisfaction: Advisory Council Members

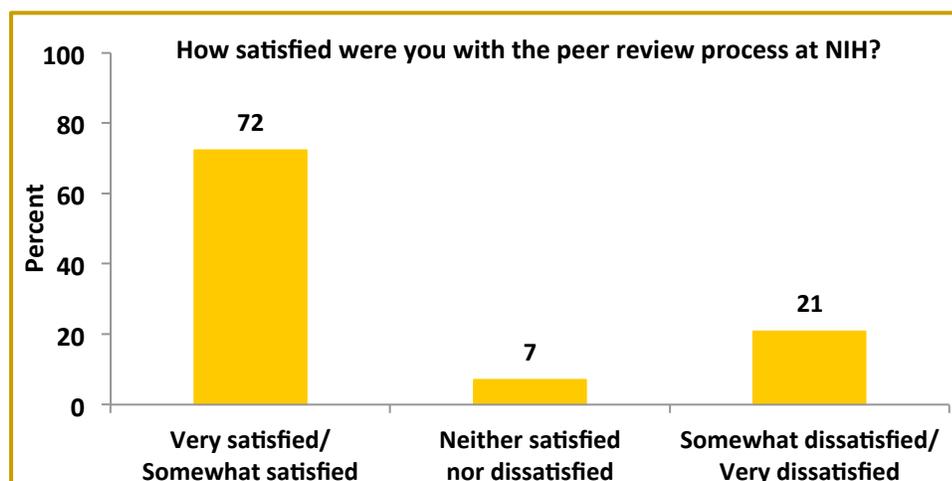
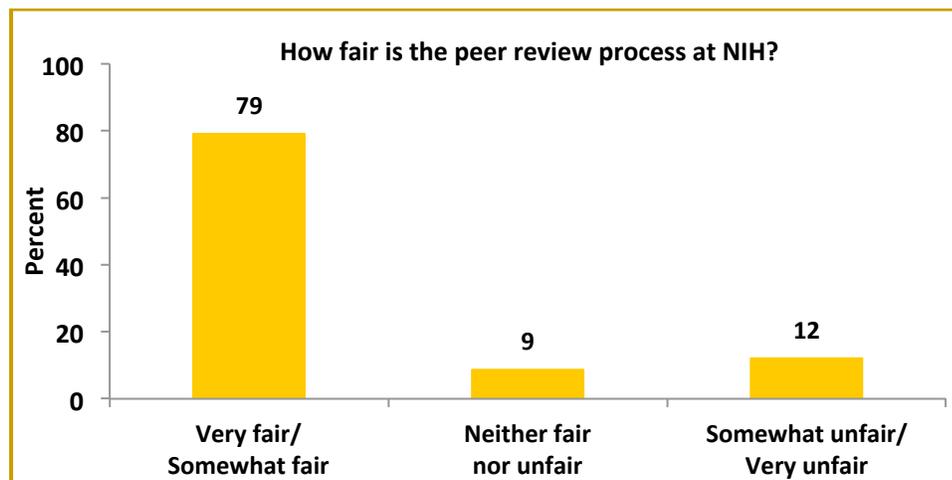
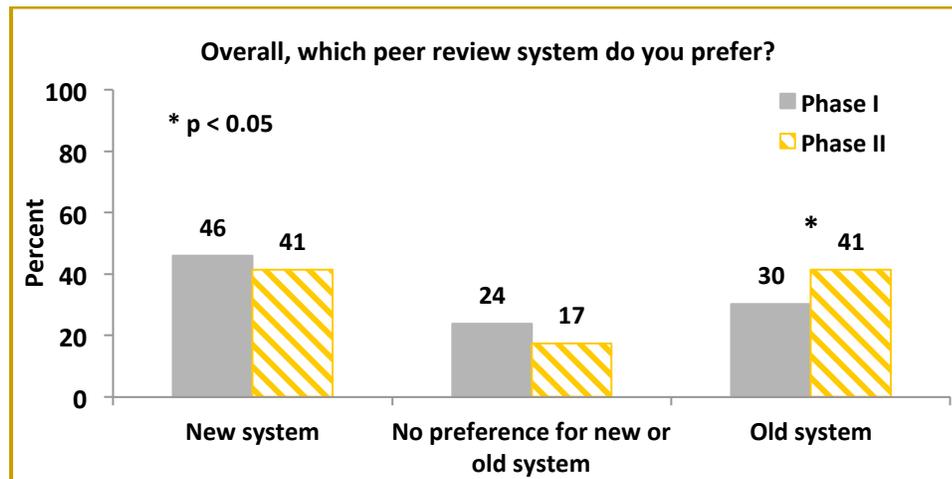


Figure 22. Advisory council member responses to the three overall evaluation questions on the Phase II surveys only. Advisory council members responded to different questions on the Phase I surveys; preferences for new vs. old peer review systems are compared in the top panel.

Overall Ratings of the Enhancing Peer Review Changes by NIH Staff Members

SROs and POs

At the end of their surveys, SROs and POs were asked to rate the extent to which each of the changes contributed positively to the objectives of the Enhancing Peer Review initiative. They were also asked to rate the extent to which further change was needed for each of these.

[Figure 23](#) shows that most of the changes were rated by at least 50% of SROs and/or POs to contribute positively to the objectives. More SROs and POs agreed than disagreed that the Narrative Overall Impact Statement, Overall Impact scores and Clustering of Applications from New/Early Stage Investigators had contributed positively to the objectives of the Enhancing Peer Review initiative.

[Figure 24](#) shows the four elements rated by at least 50% of SROs and/or POs as in need of further changes. These were the Single Resubmission Policy, the 9-point scoring scale, the critique templates and bulleted critiques, and the Continuous Submission Policy.

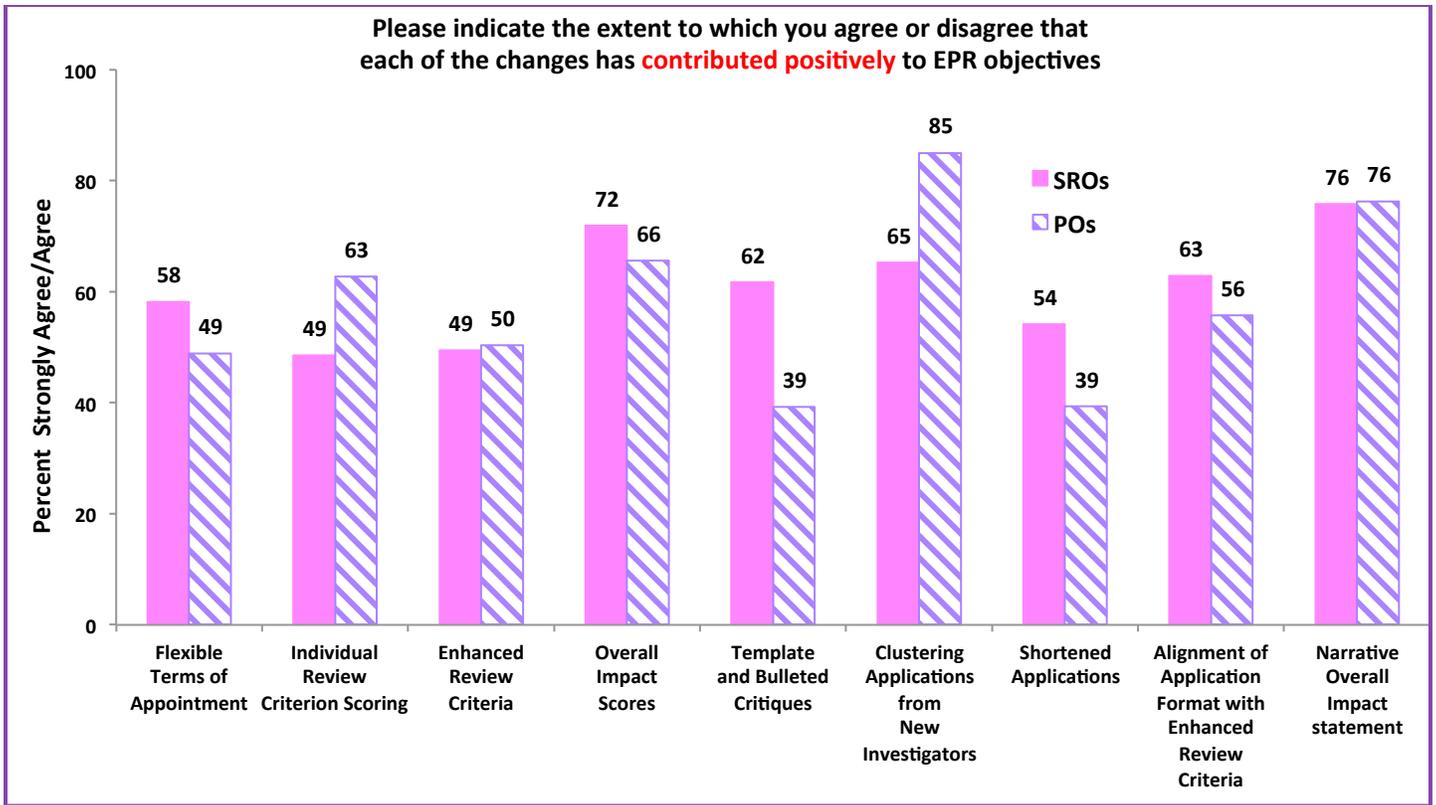


Figure 23. SRO and PO ratings when asked whether each of the Enhancing Peer Review changes had contributed positively to the objectives of the Enhancing Peer Review initiative. Statistical comparisons were not made among stakeholder groups because each group offers their opinions in the context of their unique role in the peer review process. However, it is useful to note that levels of approval of NIH staff members were similar across many of the individual changes.

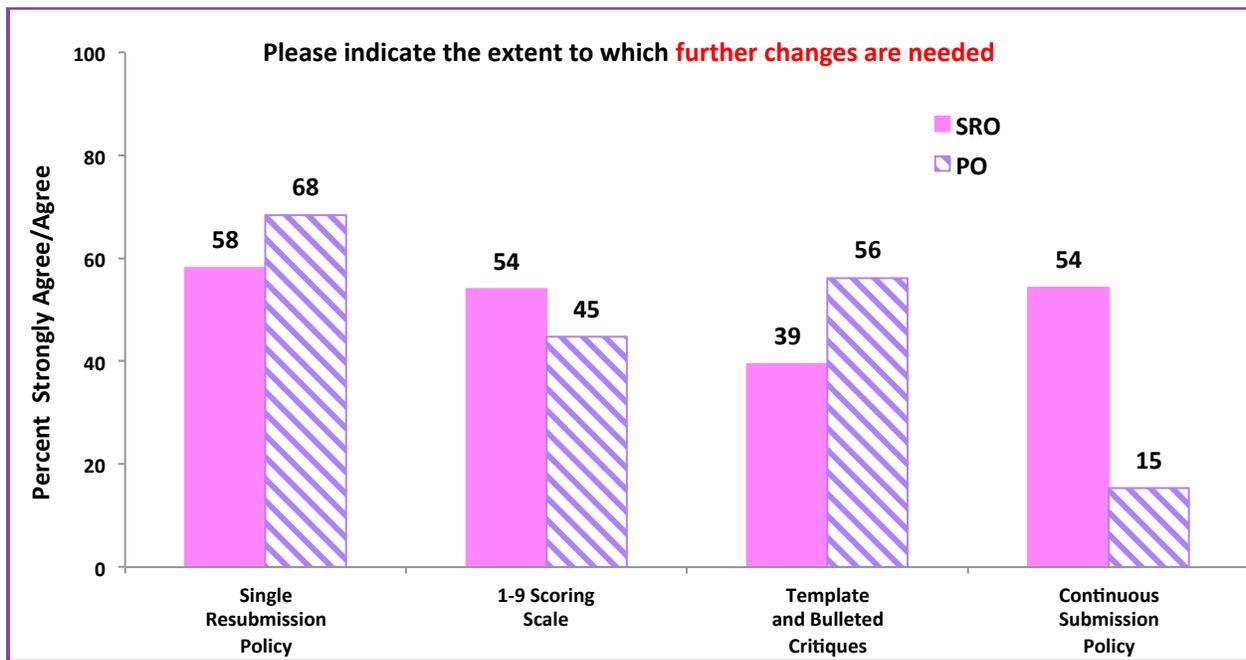


Figure 24. SRO and PO ratings when asked whether each of the Enhancing Peer Review changes was in need of further change.

Discussion

The Enhancing Peer Review initiative was crafted to clearly focus NIH's peer review system on evaluating, in a clear and concise manner, whether research proposed in NIH grant applications would exert a sustained, powerful influence on the research field(s) involved. The Enhancing Peer Review changes also aimed to improve the consistency of peer review processes and reduce the burden of review, to make it feasible for more well-qualified scientists to participate as peer reviewers.

The Enhancing Peer Review surveys have identified a number of the changes made as part of the Enhancing Peer Review initiative (or concurrently with it) that have promoted the objectives of Engaging the Best Reviewers; Improving the Quality and Transparency of Review; and Ensuring Balanced and Fair Reviews. When asked to rate each of the changes introduced as part of the Enhancing Peer Review initiative, Scientific Review Officers and Program Officers identified clustering of applications from New Investigators, the Overall Impact score, the narrative Overall Impact statement, and the aligned application format most often as the elements that contributed positively to the objectives of Enhancing Peer Review ([Figure 23](#)). POs also identified criterion scores and SROs identified the bulleted critique templates as positive contributions. The overall satisfaction ratings of NIH stakeholders also indicate general satisfaction with the peer review system, although equal proportions of POs and Advisory Council members preferred the new peer review system as preferred the old peer review system.

Enhancing Peer Review elements identified as in need of further change include the single resubmission policy, nine-point scoring system, bulleted critiques and continuous submission policy ([Figure 24](#)). Responses to structured survey questions as well as comments provided in open-ended questions on the survey provide a clearer picture of the reasons these elements of the Enhancing Peer Review initiative are identified as warranting further attention. Each of these elements will be examined to determine what refinements are feasible to address the concerns revealed in the surveys.

The nine-point scoring system was introduced to encourage reviewers to use the entire scale in evaluating applications and to address the concern that the former priority scores might be misperceived as conveying an unrealistic degree of precision. However, the responses and comments collected on the surveys indicate that reviewers still do not use the entire range of scores, and staff members are increasingly concerned about score inflation. Reviewers repeatedly expressed a desire for more scoring guidance, coaching and direct scoring instruction from SROs. NIH may be able to carry out further improvements to the scoring process to help reviewers use the scoring scale more effectively to arrive at an Overall Impact score.

The bulleted critique templates continue to receive mixed ratings. Applicants whose applications were not discussed indicated that the new summary statement format was not helpful more often than helpful for understanding why the application was not discussed ([Figure 9](#)). However, applicants who submitted a new application that was not funded agreed more often than they disagreed that the summary statements were useful for deciding whether or not to resubmit their application ([Figure 10](#)). Applicants whose applications were discussed also agreed more often than they disagreed that summary statements are useful for focusing on problem areas in the application that could be corrected ([Figure 11](#)). Applicants whose applications were not discussed disagreed with this statement more often than they agreed. The critique templates were rated as an

improvement by reviewers for ease of critique preparation. The bulleted critique format was rated by SROs to be effective for focusing reviewers' attention on the strengths and weaknesses that affect the Overall Impact of the application ([Figure 12](#)). However, SROs, POs and Advisory council members agreed less often with the statement that the bulleted comments reflect complete thoughts ([Figure 12](#), [13](#), [14](#)) than with other statements pertaining to summary statement content. In their open-ended comments, some members of all stakeholder groups gave details indicating that the bulleted comments leave open the potential for incomplete information and insufficient detail to be provided.

The single resubmission policy was introduced to reduce the time from application submission to eventual funding and to eliminate the tendency for reviewers to assess proposed research projects both on their current scientific merit and the potential for improvement if the application was revised and resubmitted. The single resubmission policy has achieved both of these objectives. However, as operationalized, the policy also eliminates the opportunity for applicants to resubmit a grant application more than once without fundamental revision. Respondents in each of the stakeholder groups indicated that the single resubmission policy hinders the peer review process ([Figures 16](#), [17](#)). Respondents in each of the stakeholder groups stated in their open-ended comments that some applicants, such as New Investigators and investigators with smaller research programs and/or limited research facilities, are disproportionately affected as a result of the policy. NIH leaders continue to monitor the policy and its implications.

The continuous submission policy was identified by 54% of SROs as in need of further change ([Figure 24](#)). In their open-ended comments, SROs indicated that the policy was disruptive and difficult to accommodate within their already stringent timelines. The continuous submission policy was introduced as a benefit to incentivize committee service on NIH's study sections and other advisory committees. As with the other peer review elements identified as in need of further changes, the continuous submission policy can be examined to assess the feasibility of refinements to address the expressed concerns.

Work is underway to address some of these issues. A working group established by the Extramural Activities Working Group will examine the foundational philosophy of the NIH scoring process to define the desired scoring paradigm. The working group then will look at current scoring patterns and scoring guidance for reviewers, and will make recommendations to correct any deficiencies found. In addition, the NIH Office of Extramural Research is conducting a needs assessment to explore more efficient ways of producing templates for reviewers' critiques. Another needs assessment is being conducted in response to recommendations from the Advisory Committee to the NIH Director Working Group on Diversity in the Biomedical Research Workforce considering the feasibility of anonymizing grant applications during peer review. Thus, the continuous review of peer review, policy development, technological developments, and implementation of review practices will continue to evolve in the coming years.

Appendix 1: The following changes were made to the NIH grant application and peer review process.

- 1) **Assessment of Overall Impact:** Beginning with applications reviewed in the May 2009 Scientific Review Group (SRG) meetings (which reviewed applications submitted for FY 2010 funding), all NIH grant applications that are discussed during the SRG meeting are now assigned a final score for Overall Impact that reflects the reviewers' assessment of the likelihood of the project to exert a sustained, powerful influence on the research field(s) involved, in consideration of the five scored review criteria. The purpose of this change is to more clearly define Overall Impact and to anchor it in the context of the NIH mission. The former measure of scientific merit was the priority score.
- 2) **Enhanced Review Criteria:** Effective with applications submitted for FY 2010 funding, the review criteria were modified to better emphasize the potential impact of the proposed research and to address the changing nature of biomedical research.
- 3) **Clustering of Applications from New and Early Stage Investigators:** Effective with applications submitted for FY 2010 funding, the order of discussion for SRG meetings was organized so that, to the extent feasible, research grant applications from New and Early Stage Investigators were discussed together to ensure that the specific considerations and review criteria for Early Stage Investigators were applied uniformly. The purpose of this Enhancing Peer Review element was to formalize a former, widely used best practice known to improve the consistency of review for applications submitted by New Investigators.
- 4) **Nine-Point Scoring Scale:** Effective with applications submitted for FY 2010 funding, the scoring scale was modified from a 1.0 to 5.0 scale with a total of 41 potential increments to a scale of 1 – 9, with a total of 9 potential increments. The purpose of this change was to encourage reviewers to use the entire scale in evaluating applications. The dual purpose was to present final Overall Impact scores as a two-digit number ranging from 10 to 90 to address the concern that the former priority scores might be misperceived as conveying an unrealistic degree of precision.
- 5) **Review Criteria Scoring:** Effective with applications submitted for FY 2010 funding, the assigned reviewers and discussants were directed to assign numeric scores of 1 – 9 to each of the five “scored” review criteria, which for research grant applications are: Significance, Investigator(s), Innovation, Approach and Environment. These criterion scores appear on all summary statements, including those for applications which were not discussed and thus not assigned a numerical score for Overall Impact.
- 6) **Critique Templates:** Effective with applications submitted for FY 2010 funding, structured critique templates were introduced to help reviewers compose critiques that are better focused on the merit of the science presented in the application and not the summative information about the application or the potential improvements that may be realized following additional rounds of review. These critique templates can be directly uploaded to the eRA Internet Assisted Review site within the NIH Commons.
- 7) **Bulleted Critique Format:** Effective with applications submitted for FY 2010 funding, the critique format for each of the scored criteria within the critique templates was modified to a bulleted list of strengths and weaknesses. The purpose of the change to a bulleted format was to discourage reviewers from summarizing the application and encourage very concise statements about the factors that contribute to or detract from the scientific merit of the application.
- 8) **Shortened Applications:** A shortened grant application was introduced in January 2010, for FY 2011 funding. The purpose of shortening the application was to shift the focus of the application to impact and uniqueness/originality, placing reduced emphasis on standard methodological details. This change was also anticipated to reduce the burden of reviewing their assigned applications so that reviewers' could read all applications assigned to the meeting. [Limits on Appendix materials](#) were established in 2007 and reminders were issued to the NIH grantee community that these limits remained in place under the shortened application format in [2010](#) and [2011](#).

- 9) **Alignment of the Application Format to the Specific Review Criteria:** Beginning with applications submitted in January 2010 for FY 2011 funding, the format of the shortened grant application was aligned with the enhanced review criteria. The purpose of this change was to ensure that reviewer and applicant expectations coincide with the scored review criteria for a more efficient and transparent application process.
- 10) **Flexible Terms for Chartered Reviewers:** The typical term for chartered Scientific Review Group (SRG) service was modified to give reviewers the choice of a 6-year term of service, with this longer term involving two SRG meetings per year instead of the traditional three SRG meetings per year required for reviewers serving a four-year term. The purpose of this change was to make it possible for more scientists, including physician scientists and other reviewers with clinical research expertise, to serve as chartered study section members. This change was introduced gradually in the Center for Scientific Review, beginning in 2010.
- 11) **Continuous Application Submission:** Reviewers who are appointed members of NIH Advisory Groups or who meet the criteria for “substantial service” are eligible for an alternate plan for submission and review of research grant applications within 120 days of submission. This alternate plan was introduced as a benefit to incentivize peer review service. This change was introduced gradually in the Center for Scientific Review and selected ICs, beginning in 2009.
- 12) **Post-submission Materials Policy:** A trans-NIH policy on acceptability of supplemental materials after the submission of the grant application was adopted. The purpose of the policy was to ensure that all applicants had the same opportunity to submit a limited but uniform set of materials for consideration by reviewers under the same conditions, and to accommodate unforeseen administrative events such as a natural disaster or loss of an investigator. The policy was made effective for all applications submitted on or after September 25, 2010.
- 13) **Single Resubmission Policy (Elimination of the Second Amended Application):** Beginning with original new applications (i.e., never submitted) and competing renewal applications submitted for the January 25, 2009 due dates (FY 2010 councils) and beyond, the NIH will accept only a single amendment to the original application. The purpose of this policy was to increase the numbers of high quality original and first amendments that could be funded earlier.
- 14) **Narrative Overall Impact Statements:** The format of the reviewers’ critique of Overall Impact was modified from bulleted to narrative format. Reviewers were instructed to write a paragraph summarizing the factors that informed their Overall Impact score. This modification was introduced in September of 2010 in response to feedback from numerous sources, including the Phase I Enhancing Peer Review surveys, about the quality of information contained in summary statements.
- 15) **Resumé Best Practices:** A set of recommendations was prepared by a working group of the Review Policy Committee, and details the purpose of the Resume and Summary of Discussion, the information that should be provided in the resume under various circumstances, and the resources needed to ensure that SROs can prepare an adequate resume. This guidance was implemented in 2011 in response to feedback collected on the Phase I Enhancing Peer Review surveys and in other venues, about the importance of the resume as a source of information for program staff to make funding recommendations as well as to advise applicants. The purpose of the guidance was to ensure that it was feasible for SROs to provide complete information in resumes.

Appendix 2. Sampling of respondents and analysis of the Enhancing Peer Review Surveys

Identifying respondent populations

All SROs, POs and Advisory council members were identified from NIH's eRA database if they were assigned to administer applications in at least one of the two review or council rounds that occurred prior to the deployment of their respective surveys (June and October 2011 review rounds, or October 2011, January 2012 council rounds). Accordingly, 430 SROs, 1,075 POs and 299 Advisory Council members were invited to complete the surveys.

The applicant population was defined as individuals drawn from NIH's eRA Commons database who submitted an R01, R03, R21, U01 or R34 application that was reviewed during the May 2011 and October 2011 NIH Advisory Councils. These council rounds were chosen to permit sufficient time for applicants to be notified of their funding decisions and/or to decide whether to resubmit the application prior to completing the survey. A total of 29,787 eligible individuals were identified in the eRA Commons database as applicants and were included in the sampling frame.

The reviewer population was defined as individuals in NIH's eRA Commons database who reviewed R01, R03, R21, U01 or R34 applications during the October 2011 or February 2012 NIH National Advisory Councils (corresponding to the June 2011 and October 2011 review rounds). The target population of reviewers includes regular/appointed/permanent and temporary/ad hoc reviewers. A total of 13,138 individuals were identified in the eRA Commons database as reviewers and were included in the sampling frame.

Some individuals belonged to both the applicant population and the reviewer population. The sampling design was developed so that these individuals were tapped only for one survey, either the Applicant or Reviewer survey..

A probability-based sampling design was created to ensure that statistical estimates would be unbiased as well as to ensure sufficient representation of various racial and ethnic groups. The number of individuals who could be contacted (not the number who were ultimately surveyed) was defined by burden limits under NIH's Office of Management and Budget (OMB) Generic Clearance No. 0925-0627. For the applicant and reviewer surveys, the total number of persons sampled under the burden limits established by the NIH guidance was 4,411. A sample of 2,484 applicants was invited to participate in the Applicant survey and a sample of 1,927 peer reviewers were invited to participate in the Reviewer survey. The 299 Advisory Council members were also counted in the OMB burden limit for this survey effort.

Survey Administration

Sampled respondents were invited to participate in the surveys using the email addresses listed in the eRA database. Tracing was performed to identify correct email addresses for individuals whose invitations "bounced" on the first attempt. Each respondent was assigned their own electronic survey to complete, and they were sent periodic reminder emails until their survey was submitted. The surveys were available to SROs and POs in November and December of 2011. The surveys were available to Applicants, Reviewers and Advisory Council members in February and March 2012.

Response Rates and Statistical Analysis

A total of 265 of the 430 SROs (62.1%), 378 of the 1,075 POs (35.2%), 882 of the 2,484 applicants (35.5%), 836 of the 1,927 reviewers (43.5%), and 175 of the 299 Advisory Council members (58.5%) responded to the survey. A non-response analysis of applicants and reviewers was conducted to determine whether the demographic characteristics of respondents differed from the individuals in the sample invited to take the survey. Summaries of the sampling frame and sample compositions, along with results of the bias analysis are shown in [Table 1](#) (applicants) and [Table 2](#) (reviewers). The bias estimates were calculated as follows:

$$(\text{weighted \# non-respondents} / \text{weighted \# sampled}) * [(\text{proportion non-respondents}) - (\text{proportion respondents})]$$

The following information collected from survey respondents was used to conduct statistical modeling of the survey responses to characterize underlying factors that may have contributed to survey responses:

SROs and POs:

- Number of years in their current position (as an SRO or PO)
- Number of applications (SROs) or grants (POs) typically assigned to them
- SROs: Whether respondent works in an Institute or Center or in the Center for Scientific Review
- POs: How many study sections meetings they typically attend in a round by phone or in person

Applicants and Reviewers:

- Age
- Gender
- Race
- Ethnicity
- Applicants: New Investigator status
- Applicants: Whether the application proposed a clinical research project
- Applicants: Whether the application was a resubmission
- Applicants: Whether the application was assigned and Overall Impact score
- Applicants: Whether the application was funded
- Reviewers: Earliest year of review service
- Reviewers: Whether the reviewer has completed a term of chartered service

Applicant Response Bias Estimates

Gender	Number in Sampling Frame	Proportion in Sampling Frame	Sampled	Respondents	Bias Estimate
Female	9,382	0.31	0.32	0.37	0.041
Male	19,564	0.66	0.64	0.60	-0.038
Degree	Number in Sampling Frame	Proportion in Sampling Frame	Sampled	Respondents	Bias Estimate
MD	4,353	0.15	0.14	0.11	-0.027
MD-PhD	2,957	0.10	0.09	0.08	-0.005
PhD	20,392	0.68	0.71	0.77	0.057
Age Group	Number in Sampling Frame	Proportion in Sampling Frame	Sampled	Respondents	Bias Estimate
Under 35	4,042	0.14	0.14	0.13	-0.010
35 to 40	4,612	0.15	0.16	0.15	-0.009
41 to 45	5,166	0.17	0.18	0.19	0.009
46 to 50	5,048	0.17	0.15	0.14	-0.015
51 to 55	4,071	0.14	0.13	0.13	-0.005
56 to 60	3,230	0.11	0.10	0.13	0.023
61 to 65	2,096	0.07	0.07	0.08	0.007
66 to 70	1,010	0.03	0.03	0.02	-0.012
Over 70	512	0.02	0.02	0.03	0.003
Awarded	Number in Sampling Frame	Proportion in Sampling Frame	Sampled	Respondents	Bias Estimate
No	24,334	0.82	0.82	0.78	-0.040
Yes	5,453	0.18	0.18	0.22	0.034
Race	Number in Sampling Frame	Proportion in Sampling Frame	Sampled	Respondents	Bias Estimate
Asian	5,511	0.19	0.19	0.19	0.003
Black	503	0.02	0.02	0.01	-0.003
Multiple Races	254	0.01	0.01	0.01	0.0003
Unknown	5,463	0.18	0.19	0.14	-0.061
White	17,988	0.60	0.60	0.65	0.050
Ethnicity	Number in Sampling Frame	Proportion in Sampling Frame	Sampled	Respondents	Bias Estimate
Hispanic	1,065	0.04	0.04	0.03	-0.001
Non-Hispanic	18,065	0.61	0.61	0.66	0.042
Unknown	9,881	0.33	0.33	0.28	-0.051

Table 1. Applicant characteristics (in the first column) stored in their eRA Commons person profile, and the award status of the application which rendered them eligible to participate in the Enhancing Peer Review survey, at the time the sampling frame was constructed. The total number of applicants and the proportion of applicants in each group is indicated in the second and third columns. The proportion of each group sampled and proportion responding to the survey are indicated in columns 4 and 5, respectively. Applicants whose person profiles contained missing information were included in the sampling frame, but the counts are omitted from the table. Bias estimates were not calculated for characteristics which represented less than 1% of the sampling frame composition.

Reviewer Response Bias Estimates

Gender	Number in Sampling Frame	Proportion in Sampling Frame	Sampled	Respondents	Bias Estimate
Female	4,253	0.32	0.33	0.38	0.044
Male	8,778	0.67	0.66	0.61	-0.053
Degree	Number in Sampling Frame	Proportion in Sampling Frame	Sampled	Respondents	Bias Estimate
MD	2,306	0.18	0.17	0.13	-0.046
MD-PhD	1,340	0.10	0.10	0.09	-0.013
PhD	9,362	0.71	0.71	0.77	0.051
Age Group	Number in Sampling Frame	Proportion in Sampling Frame	Sampled	Respondents	Bias Estimate
Under 35	1,224	0.09	0.09	0.08	-0.007
35 to 40	972	0.07	0.07	0.07	-0.006
41 to 45	1,969	0.15	0.15	0.16	0.007
46 to 50	2,537	0.19	0.17	0.17	0.006
51 to 55	2,430	0.18	0.21	0.22	0.005
56 to 60	1,986	0.15	0.16	0.17	0.006
61 to 65	1,199	0.09	0.08	0.08	0.0005
66 to 70	596	0.05	0.05	0.04	-0.004
Over 70	225	0.02	0.02	0.01	-0.006
Awarded	Number in Sampling Frame	Proportion in Sampling Frame	Sampled	Respondents	Bias Estimate
No	11,551	0.88	0.88	0.88	-0.0001
Yes	1,587	0.12	0.12	0.12	0.0001
Race	Number in Sampling Frame	Proportion in Sampling Frame	Sampled	Respondents	Bias Estimate
Asian	1,988	0.15	0.15	0.14	-0.008
Black	276	0.02	0.02	0.02	-0.003
Multiple Races	134	0.01	0.01	0.01	-0.003
Unknown	1,058	0.08	0.09	0.07	-0.024
White	9,647	0.73	0.73	0.76	0.031
Ethnicity	Number in Sampling Frame	Proportion in Sampling Frame	Sampled	Respondents	Bias Estimate
Hispanic	580	0.04	0.04	0.05	0.002
Non-Hispanic	8,966	0.68	0.69	0.74	0.043
Unknown	3,284	0.25	0.24	0.19	-0.053

Table 2. Reviewer characteristics (in the first column) indicated in their eRA Commons person profile, and the award status of the application which rendered them eligible to participate in the Enhancing Peer Review survey, at the time the sampling frame was constructed. The proportion of each group sampled and proportion responding to the survey are indicated in columns 4 and 5, respectively. The proportion of each group sampled and the proportion responding to the survey are indicated in columns 4 and 5. Reviewers whose person profiles contained missing information were included in the sampling frame, but the counts are omitted from the table. Bias estimates were not calculated for characteristics which represented less than 1% of the sampling frame composition.

Appendix 3. Acknowledgements

The following NIH staff members contributed to the development and/or analysis of the Phase 2 Enhancing Peer Review surveys.

Neeraj Agarwal, Ph.D.
National Eye Institute

Rashada Alexander, Ph.D.
Office of Extramural Research
Office of the Director

Sally Amero, Ph.D.
Office of Extramural Research
Office of the Director

Regina Bures, Ph.D.
Eunice Kennedy Shriver National Institute
for Child Health and Human Development

George Chacko, Ph.D.
Center for Scientific Review

Mark Egli, Ph.D.
National Institute on Alcohol Abuse and
Alcoholism

Della Hann, Ph.D.
Office of Extramural Research
Office of the Director

Meenaxi Hiremath, Ph.D.
National Institute for Drug Abuse

Zoe Huang, M.D.
National Library of Medicine

Cheryl Kitt, Ph.D.
Center for Scientific Review

Helen Lin, Ph.D.
National Institute of Arthritis,
Musculoskeletal and Skin Diseases

Peter Lyster, Ph.D.
National Institute of General Medical
Sciences

Linda Piccinino, M.P.S.
Division of Program Coordination, Planning
and Strategic Initiatives
Office of the Director

Richard Rippe, Ph.D.
National Institute on Alcohol Abuse and
Alcoholism

Luci Roberts, Ph.D.
Office of Extramural Research
Office of the Director

Marvin Salin, Ph.D.
National Cancer Institute

Cary Scheiderer, Ph.D.
Office of Extramural Research
Office of the Director

Aileen Schulte, Ph.D.
National Institute on Mental Health

Chelvi Thyagarajan, Ph.D.
National Institute on Nursing Research

Thomas Vollberg, Ph.D.
National Cancer Institute

Cheri Wiggs, Ph.D.
National Eye Institute

David Wilson, Ph.D.
National Heart, Lung and Blood Institute