Office of Extramural Research

Enhancing Peer Review
Survey Results Report

Table of Contents

Executive Summary

The National Institutes of Health (NIH) has a longstanding history of supporting the most promising and meritorious biomedical and behavioral research using a broad range of approaches, strategies and mechanisms. In 2007, NIH recognized that as the scientific and public health landscape evolved, it was critical that the processes used to support science continue to be fair, efficient, and effective. Thus, the NIH embarked on a self-study[1] of its peer review process. Recommendations arising from the peer review self-study resulted in specific changes that comprise the Enhancing Peer Review Initiative.

This report summarizes the results of five stakeholder satisfaction surveys, sponsored by NIH, which examined Enhancing Peer Review changes introduced in May 2009:

1) a nine-point scoring system
2) criterion scores
3) a bulleted critique format and structured critique templates
4) enhanced review criteria
5) clustering of applications that propose clinical research and clustering of R01 applications submitted by New Investigators

NIH administered the surveys to five stakeholder groups, including NIH grant applicants, NIH peer reviewers, Scientific Review Officers (SROs), Program Officers (POs) and Advisory Council members. Grant applicants and reviewers were asked to rate their agreement or disagreement with statements about the changes to the peer review system introduced in May 2009, or the former peer review system, based upon the timing of their most recent interaction with the NIH peer review process. SROs, POs, and Advisory Council members were asked to rate their current opinions about the peer review system after the changes were introduced in May 2009.

Results of the Enhancing Peer Review Surveys

Nine-point Scoring System: The survey of reviewers indicated that the nine-point scoring range was adequate for reviewers to communicate meaningful differences in the quality of applications. Advisory Council members indicated that the nine-point scale was easy for them to understand.

Criterion Scores: Forty-nine percent of reviewers "strongly agreed/agreed" that criterion scores were helpful to them for communicating why an application was not discussed. Additionally, program officers rated criterion scores as one of the enhancing peer review changes that had been most helpful for advising applicants after review. However, applicants' responses reflected only moderate agreement about whether criterion scores were helpful for understanding an application's strengths and weaknesses or the problem areas that could be corrected. These results suggest that the NIH may want to explore ways to improve the helpfulness of criterion scores for all stakeholders.

---

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

Executive Summary (continued)

Bulleted Critiques: Applicants agreed that both the new, bulleted summary statement format and the old, narrative summary statement format helped them to focus on problem areas that could be corrected. However, reviewers who rated the peer review changes introduced in May 2009 were significantly less likely to "strongly agree/agree" than reviewers who rated the former peer review system that the critique format was adequate for capturing the strengths and weaknesses of the application. SROs responses reflected no clear agreement about whether the bulleted format was useful for focusing the critiques on factors that influence final score. POs "disagreed/ strongly disagreed" more often than they "strongly agreed/agreed" that summary statements in the new structured format were helpful for explaining the recommendations of the review group.

The Enhancing Peer Review surveys also examined whether bulleted critiques effectively communicated the reason(s) applications were not discussed. Applicants, SROs and POs "disagreed/strongly disagreed" more often than they "strongly agreed/agreed" that the new summary statement format helped them to understand why applications were not discussed. Reviewers who were asked to rate the new bulleted critique format agreed significantly less often and disagreed significantly more often that the format was helpful for communicating to applicants why their applications were not discussed.

Reviewers "strongly agreed/agreed" more often than they "disagreed/strongly disagreed" that both formats were helpful to them for completing their critiques efficiently. However, reviewers who rated the new structured critique templates "disagreed/strongly disagreed" significantly less often that the format was helpful in completing their critiques efficiently in comparison to reviewers who rated the old, narrative format.

Taken together, these results suggest that the new structured critique templates have benefitted reviewers in terms of efficiency. However, few stakeholders rated the bulleted critique format as helpful for communicating information about the pertinent factors that affected the outcome of the review. These results suggest that the NIH may want to explore ways to improve the helpfulness of bulleted critiques for all stakeholders.

Enhanced Review Criteria: POs and SROs were more likely to strongly disagree than to strongly agree that the enhanced review criteria resulted in greater clarity regarding the strengths and weaknesses of the application.

Clustering of Clinical Research and New Investigator Applications: Most reviewers "strongly agreed/agreed" that clustering resulted in a more consistent review of affected applications. POs and SROs selected clustering most frequently as a change that had contributed positively to the Enhancing Peer Review objectives.

Preference for New Peer Review System versus Old: Applicants who were experienced with the new peer review system expressed no significant preference for the new system versus the old. However, most reviewers who were asked to rate the peer review changes introduced in May 2009 expressed a preference for the new peer review system over the old system. SROs expressed no clear preference for the old peer review system versus the new peer review system, whereas POs slightly preferred the old peer review system over the new system. Finally, Advisory Council members expressed a moderate preference for the new system over the old.

Executive Summary (continued)

Overall Fairness and Satisfaction: The pattern of results for applicants who rated the peer review system before and after the changes was similar. Applicants rated the peer review system as fair or very fair most often and rated themselves as satisfied or very satisfied.

Although reviewers rated the peer review system after the changes were introduced as "very fair/fair" significantly less often, and rated themselves as "very satisfied/satisfied" with the peer review system significantly less often after the changes were introduced than reviewers who rated the old peer review system, reviewers in both groups reported high levels of fairness and satisfaction.

Most SROs responded that the peer review system remains very fair or fair after the Enhancing Peer Review changes. Approximately equal numbers of SROs rated themselves as "very satisfied/satisfied" as rated themselves "dissatisfied/very dissatisfied." POs' responses to questions about fairness and satisfaction were very similar to those for SROs: they responded that the new peer review system was "very fair/fair", although approximately equal numbers of POs rated themselves as "very satisfied/satisfied" as rated themselves "dissatisfied/very dissatisfied."

Fifty percent of Advisory Council members indicated that their perceptions of the fairness of the peer review system had not changed since the introduction of the Enhancing Peer Review changes and approximately equal numbers rated their own satisfaction as having improved, remained the same or worsened since the Enhancing Peer Review changes were introduced.

Continuous Review of Peer Review: The Enhancing Peer Review Initiative was conceived as an ongoing process of refinement to ensure that the NIH peer review system continues to evolve as science evolves. Continual monitoring and assessment are needed to facilitate this process. The Enhancing Peer Review surveys presented here were an early snapshot of the opinions of NIH stakeholders about the changes to the peer review process that were implemented in May 2009. Only one complete application cycle had occurred when the applicant and reviewer surveys were deployed in December 2009, and NIH staff members had administered two full application cycles prior to being surveyed in April 2010.

NIH has been continuously monitoring feedback on the peer review changes since their inception and has already implemented a number of refinements to the new peer review system in response to this information. Most notably, in January 2010, additional guidance was issued to clarify distinctions between Overall Impact and the review criterion Significance; in September 2010 NIH began requiring reviewers to include a narrative statement to explain the overall impact score. NIH will reassess stakeholder opinions of the scoring system, the critique format and other peer review processes at a later date. In addition, NIH's ongoing review of the NIH peer review system will also examine the shortened applications and alignment of research plans with the NIH review criteria.

Introduction

NIH Implemented the Enhancing Peer Review Initiative in three stages.

Stage 1 (January 2009):
- Elimination of A2 applications announced
- Early stage investigator policy announced

Stage 2 (May 2009) Changes to the Peer Review System:
- Nine-point scoring system
- Criterion scores
- Bulleted critiques
- Enhanced review criteria
- Clustering of clinical and New Investigator applications

Stage 3 (January 2010) Changes to the NIH Grant Application Introduced:
- Alignment of applications with criteria
- Shortened research strategy
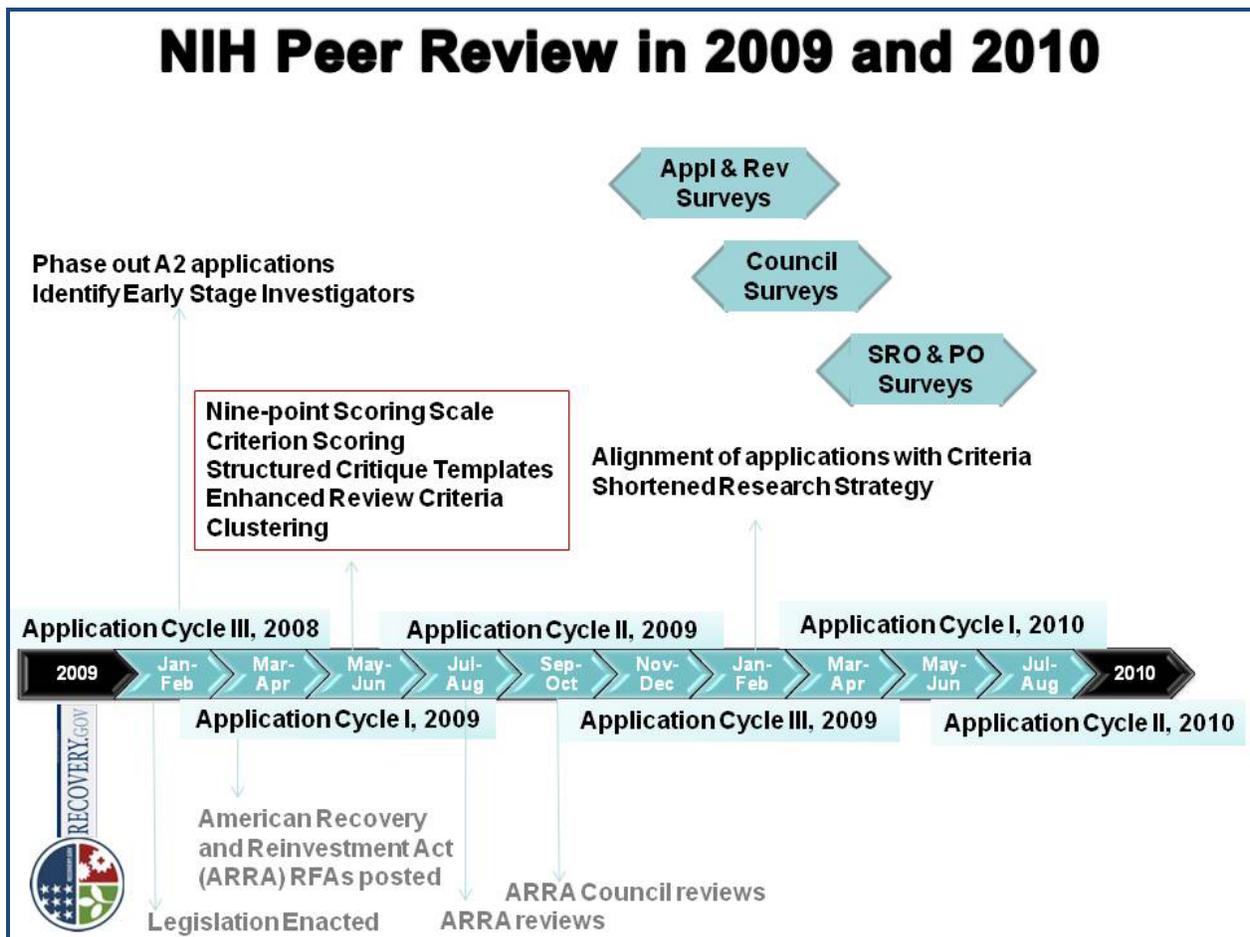
The Surveys

- The goal of the current surveys was to assess the perceptions of stakeholders about the **Stage 2** changes only.

- The surveys were conducted very early after the changes were introduced to enable the before-after design of the applicant and reviewer surveys.

The Applicant and Reviewer survey data were analyzed using multinomial modeling for categorical data.
The following variables were included in the model:

- Experienced with peer review changes vs. not experienced
- Whether an applicant's most recent application was funded
- Whether reviewers have experience as applicants
- Gender
- Age (45 and younger; 46 and older)
- Professional rank (full professor; associate professor/senior scientist; or Other)
- Institution type (university versus other)
- Education (Ph.D. versus other)
- Year first NIH Grant application was submitted (1990 or earlier; 1991-present)

The Enhancing Peer Review changes found to be statistically significant are presented in this report as predicted probabilities, taking into account all of the other variables listed above.

Number of Respondents:

374 Applicants rated peer review system BEFORE changes;
    504 Applicants rated peer review system AFTER the changes.

221 Reviewers rated the peer review system BEFORE changes;
    537 Reviewers rated peer review system AFTER the changes.

288 SROs and **437 POs** rated the system AFTER the changes.

291 Advisory Council members rated the system AFTER the changes.

http://enhancing-peer-review.nih.gov/

Nine-point Scoring Scale

- The previous scoring system offered reviewers a range of scores from 1.0 to 5.0, for a total of 41 possible scores that could be assigned to applications.

- There was concern that it was not possible for a reviewer to discriminate application quality to this level of precision.

- The nine-point scoring scale was implemented to address this concern by reducing the number of increments available to reviewers to 9.

- The details of this decision were communicated to the scientific community in http://grants.nih.gov/grants/guide/notice-files/NOT-OD-09-024.html.

Results of survey questions assessing the adequacy of the nine-point scoring scale:

Reviewers:

- Most Reviewers agreed/strongly agreed that the nine-point scoring scale had sufficient range to communicate meaningful differences in the quality of the applications.

- Reviewers agreed that both the old (1.0 - 5.0) and new nine-point scoring scales had sufficient range.

Advisory Council Members:

- Most Advisory Council members also agreed or strongly agreed that the nine-point scoring scale was easy for them to understand.

Nine-point Scoring Scale (continued)

Reviewers: The [1-5 / 1-9] scale had sufficient range for me to communicate meaningful differences in the quality of the applications.

■ No Experience with Enhancing Peer Review Changes
■ Experienced with Enhancing Peer Review Changes

	Agree/Strongly agree	Neither agree nor disagree	Disagree/Strongly disagree
No Experience	85	5	11
Experienced	79	5	17

Advisory Council Members: It is easy for me to understand the 1-9 scoring scale (translates to 10-90 score range).

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
20	47	16	13	4

Criterion Scores

- Criterion scores were developed to improve transparency by providing a means to communicate ratings from assigned reviewers even when the application is not discussed.

- Before the review meeting, each reviewer and discussant assigned to an application is asked to give a separate score for each of five review criteria (Significance, Investigators, Innovation, Approach, and Environment).   These scores are reported individually on the summary statement.

The survey questions assessing the criterion scoring examined two aspects of criterion scores: their perceived <u>helpfulness</u> and the <u>consistency</u> of the scores.

Results for questions about criterion scoring include the following:

Helpfulness

Applicants:

There was moderate agreement that criterion scores were helpful for understanding:

- The strengths and weaknesses of the application, or
- The problem areas that could be corrected.

Reviewers:

- Forty-nine percent of reviewers strongly agreed or agreed that criterion scores were helpful to them in communicating why an application was not discussed.

POs:

- Program officers rated criterion scores as one of the Enhancing Peer Review changes that had been most helpful for advising applicants after review.

Criterion Scores (continued)

Applicants: The individual criterion scores were helpful in understanding the strengths and weaknesses of the application.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
6	35	26	23	11

Applicants: The individual criterion scores helped me focus on problem areas in the application that could be corrected.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
9	36	22	23	11

Criterion Scores (continued)

Reviewers: The criterion scores helped me communicate to the applicants why their applications were not discussed.

Response	Percent
Strongly agree	9
Agree	40
Neither agree nor disagree	26
Disagree	17
Strongly disagree	8

POs: Enhancing Peer Review changes that have been *most* helpful for advising applicants after review.

Change	Percent
Nine-point scoring	15
Enhanced review criteria	12
Reviewer scores for each review criterion	45
Overall Impact/Priority Score	23
Use of template for critiques	16
Bulleted comments	22
Clustering of applications from New Investigators	18
None of the changes have been helpful	22

Consistency

SROs and POs disagreed/strongly disagreed more often than they agreed when asked if criterion scores were consistent with overall impact/priority scores.

SROs: Final criterion scores are consistent with overall impact/priority scores

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
1	29	27	28	15

POs: The overall impact/priority score is generally consistent with the individual criterion scores.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
2	26	27	32	13
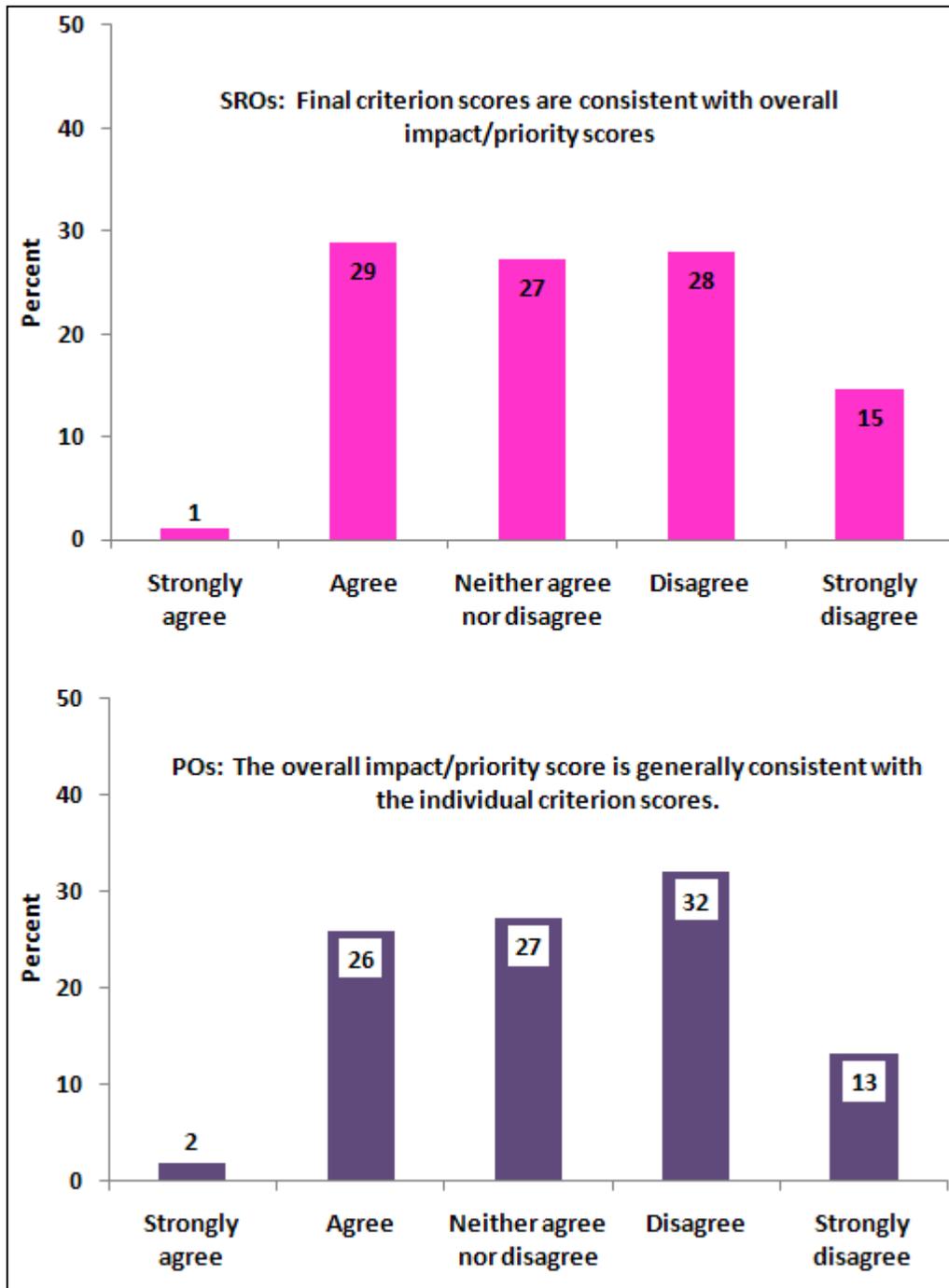
Bulleted Comments and Structured Critique Templates

- The format of critiques was standardized using fillable critique templates to streamline the development of critiques.

- The former narrative critique format was replaced by a bulleted list of the strengths and weaknesses for each of the review criteria.

- The purpose of the bulleted comments was to help focus the review on the factors that influence the merit of the application.

Results of the following questions addressed whether the bulleted comments helped focus the critiques on strengths and weaknesses:

Applicants:

- Applicants agreed that both old and new summary statement formats helped them to focus on problem areas that could be corrected.

Reviewers:

- Reviewers who rated the narrative critique format agreed significantly more often than those who rated the bulleted critique format that the bulleted format is adequate for capturing strengths and weaknesses.

SROs:

- There was no clear agreement among SROs about whether the bulleted critique format is useful for focusing the critiques on factors that influence scores.

POs:

- POs disagreed/strongly disagreed more often than they agreed/strongly agreed that the bulleted comments were helpful for explaining the recommendations of the review group.

Bulleted Comments and Structured Critique Templates (continued)

Applicants: The Summary Statement helped me focus on problem areas in the application that could be corrected.

- No Experience with Enhancing Peer Review Changes
- Experienced with Enhancing Peer Review Changes

	Agree/Strongly agree	Neither agree nor disagree	Disagree/Strongly disagree
No Experience	67	20	14
Experienced	68	16	16

Reviewers: The [narrative / bulleted] critique format was adequate for capturing the strengths and weaknesses.

* p < 0.05

- No Experience with Enhancing Peer Review Changes
- Experienced with Enhancing Peer Review Changes

	Agree/Strongly agree	Neither agree nor disagree	Disagree/Strongly disagree
No Experience	93	3	4
Experienced	67 *	10 *	23 *

SROs: Bulleted critique format useful for focusing on factors that influence score.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
8	37	24	16	15

POs: Summary statements are helpful for explaining the recommendations of the review group.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
5	25	12	31	27

http://enhancing-peer-review.nih.gov/

Bulleted Comments and Structured Critique Templates (continued)

Results of the following questions addressed whether the bulleted comments provided the information needed to understand why the applications were not discussed:


Applicants, SROs and POs:

- Applicants, SROs and POs all disagreed/strongly disagreed more often than they strongly agreed/agreed that the summary statements containing bulleted critiques, helped them to understand why applications were not discussed.

Reviewers:

- Reviewers who rated the narrative critique format agreed/strongly agreed significantly more often than reviewers who rated the bulleted format that it was helpful for communicating to applicants why their applications were not discussed.

## Bulleted Comments and Structured Critique Templates (continued)

Applicants: The summary statement helped me understand why the scientific review group did not discuss my application.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
3	30	12	36	19

Reviewers: The [narrative/ bulleted] critique format helped me communicate to applicants why their applications were not discussed.

- No Experience with Enhancing Peer Review Changes
- Experienced with Enhancing Peer Review Changes

	Agree/ Strongly agree	Neither agree nor disagree	Disagree/ Strongly disagree
No Experience	69 *	15	16 *
Experienced	46	19	35

* p < 0.05

SROs: Bulleted critiques are helpful for understanding why applications were not discussed.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
2	24	33	21	19

POs: Critiques are helpful for advising applicants about the reasons their application was not discussed.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
2	26	24	27	21

http://enhancing-peer-review.nih.gov/

Bulleted Comments and Structured Critique Templates (continued)

Results of the following question addressed whether the structured templates helped to streamline the preparation of critiques:

Reviewers:

- Reviewers strongly agreed/agreed more often than they disagreed/strongly disagreed that both the narrative and bulleted critique formats helped them complete their critiques efficiently.

- Reviewers who rated the bulleted critique format disagreed/strongly disagreed with this statement significantly less often than those who rated the narrative format.

Reviewers: The [narrative critique format/ structured critique templates] helped me complete my critiques efficiently.

- No Experience with Enhancing Peer Review Changes
- Experienced with Enhancing Peer Review Changes

	Agree/Strongly agree	Neither agree nor disagree	Disagree/Strongly disagree
No Experience	66	17	17
Experienced	76	16	9 *

* p < 0.05
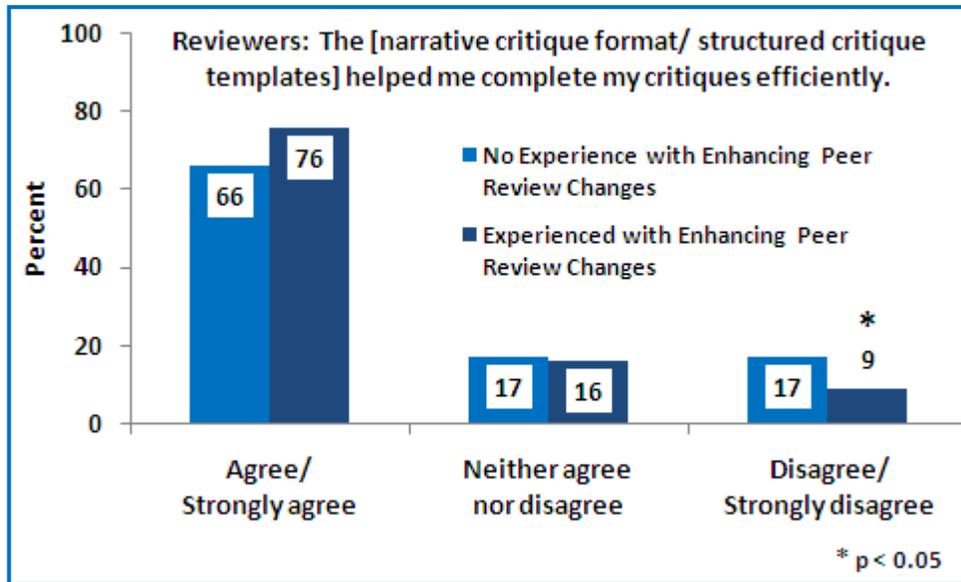
Enhanced Review Criteria

The review criteria were modified as follows:

- Scientific merit was defined as the overall impact on the research field(s) involved.
- More emphasis was placed on Investigator(s), and less was placed on Approach.
- Guidelines for evaluating merit for New and Early Stage Investigators were better developed in Investigator(s).
- Guidelines for reviewing clinical research studies and basic technology development were incorporated into Approach.
- Innovation was redefined to address both paradigm shifts and tests of feasibility.

SROs and POs

- Results of the following question addressed whether the enhanced review criteria result in greater clarity regarding strengths and weaknesses of the application.
- Both were more likely to strongly disagree than strongly agree that the enhanced review criteria result in greater clarity regarding strengths and weaknesses of the application.

The enhanced review criteria result in greater clarity regarding strengths and weaknesses of the application.
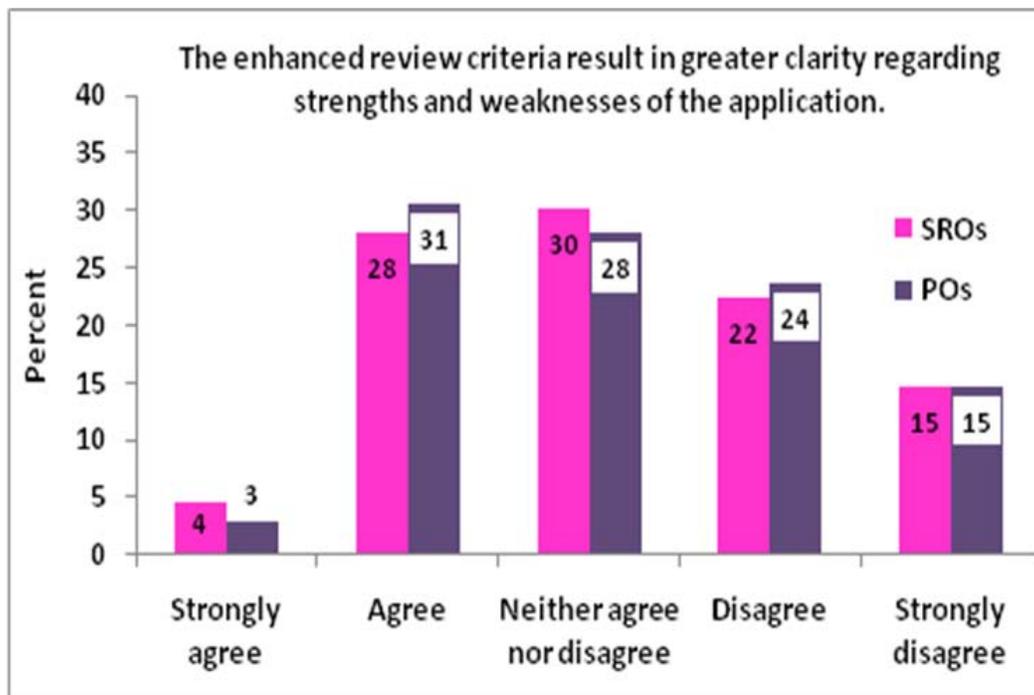
Clustering of New Investigator and Clinical Applications

Clustering is the process of grouping applications that share a similar attribute within the order of review for a study section meeting, so they are deliberated in succession.

- Clustering promotes improved attention by reviewers to unique review criteria for applications submitted by New Investigators, applications submitted under non-traditional grants activities, and for applications that share a similar attribute that must be attended to during discussions, such as applications that propose clinical research.

- The Enhancing Peer Review Enhancements Initiative formalized the practice of clustering, where feasible, for R01 applications submitted by New Investigators and for grant applications involving human subjects that are reviewed in "mixed" study sections, i.e., study sections that see clinical and non-clinical applications.

Results for clustering of New Investigator and clinical research applications:

Reviewers:

- Reviewers were more likely to strongly agree/agree than disagree/strongly disagree that clustering of New Investigator and clinical research applications resulted in a more consistent review of these applications.

SROs and POs:

- Out of a number of changes, SROs and POs both selected clustering most frequently as a change that had contributed positively to the Enhancing Peer Review objectives.

http://enhancing-peer-review.nih.gov/

Clustering of New Investigator and Clinical Applications (continued)

Reviewers: Clustering applications from New Investigators resulted in a more consistent review of those applications.

Response	Percent
Strongly agree	22
Agree	48
Neither agree nor disagree	20
Disagree	9
Strongly disagree	1

Reviewers: Clustering clinical applications resulted in a more consistent review of those applications.

Response	Percent
Strongly agree	12
Agree	42
Neither agree nor disagree	35
Disagree	10
Strongly disagree	1

To what extent do you agree that each of the following has contributed *positively* to the objectives of the Enhancing Peer Review Initiative (strongly agree/agree).

	SROs	POs
Flexible Terms of Appointment	19	
Nine-point Scoring Scale	26	37
Individual Review Criterion Scoring	38	53
Enhanced Review Criteria	27	42
Overall Impact /Priority Scores	39	37
Template and Bulleted Critiques	34	29
Clustering Applications from New Investigators	47	72

Overall Satisfaction: Applicants and Reviewers

Applicants

- Applicants whose applications were reviewed after the Enhancing Peer Review changes were introduced expressed no preference for the new system over the old.

- The pattern of results for applicants who rated the peer review system before and after the changes was similar. Applicants rated the peer review system as fair or very fair most often and rated themselves as satisfied or very satisfied.

Reviewers

- Reviewers who served on study sections under the new peer review system expressed a preference for the new system over the old system.

- Reviewers who were asked about the old peer review system rated it very fair/ fair significantly more often than reviewers who were asked about the new system.

- Reviewers who were asked about the old system also indicated they were very satisfied/satisfied significantly more often than reviewers who were asked about the new system.

http://enhancing-peer-review.nih.gov/

Overall Satisfaction: Applicants and Reviewers (continued)

Applicant and Reviewer: Which Peer Review System do you prefer?

- Applied after Enhancements
- Reviewed after Enhancements

	New system	Old system	No preference
Applied after Enhancements	39	38	24
Reviewed after Enhancements	60	25	15

How fair is the peer review process at NIH [prior to/ after the] Enhancing Peer Review changes were introduced in May 2009?

* p < 0.05

	Very fair/ Somewhat fair	Neither fair nor unfair	Somewhat unfair/ Very unfair
Applied before Enhancements	71	14	16
Applied after Enhancements	58	22	20
Reviewed before Enhancements	87 *	6	7 *
Reviewed after Enhancements	74	10	16

- Applied before Enhancements
- Applied after Enhancements
- Reviewed before Enhancements
- Reviewed after Enhancements

How satisfied are you with the peer review process at NIH [prior to/ after the] Enhancing Peer Review changes were introduced in May 2009?

	Very satisfied/ Somewhat satisfied	Neither satisfied nor dissatisfied	Somewhat dissatisfied/ Very dissatisfied
Applied before Enhancements	66	10	24
Applied after Enhancements	53	16	32
Reviewed before Enhancements	78 *	9	13 *
Reviewed after Enhancements	66	10	24

Overall Satisfaction: SROs, POs and Advisory Council Members

SROs, POs and Advisory Council Members' Preferences:

- SROs expressed no clear preference for the old peer review system versus the new.

- POs slightly preferred the old system over the new.

- Advisory Council members slightly preferred the new system over the old.

SROs and POs' Ratings of Fairness and Satisfaction:

- Most SROs and POs responded that the peer review system remains very fair/fair after the Enhancing Peer Review changes.

- Approximately equal numbers of SROs and POs were very satisfied/satisfied as were dissatisfied/very dissatisfied.

Overall Satisfaction: SROs, POs and Advisory Council Members
(continued)

NIH staff: Which peer review system do you prefer?

	SRO	PO	Advisory Council
New system	35	27	46
Old system	34	45	30
No preference	31	28	24

How fair is the peer review process at NIH since the Enhancing Peer Review changes were introduced in May 2009?

	SRO	PO
Very fair/Somewhat fair	66	58
Neither fair nor unfair	20	25
Somewhat unfair/Very unfair	15	18

How satisfied are you with the peer review process at NIH since the Enhancing Peer Review changes were introduced in May 2009?

	SRO	PO
Very satisfied/Somewhat satisfied	46	40
Neither satisfied nor dissatisfied	19	19
Somewhat dissatisfied/Very dissatisfied	35	43

http://enhancing-peer-review.nih.gov/

Overall Satisfaction: SROs, POs and Advisory Council Members
(continued)

Advisory Council Members:

- Fifty percent of Advisory Council members responded that the fairness of the NIH peer review system had not changed as a result of the Enhancing Peer Review changes.

- They were approximately equally likely to rate their own satisfaction as having improved, remained the same, or worsened since the Enhancing Peer Review changes were introduced.
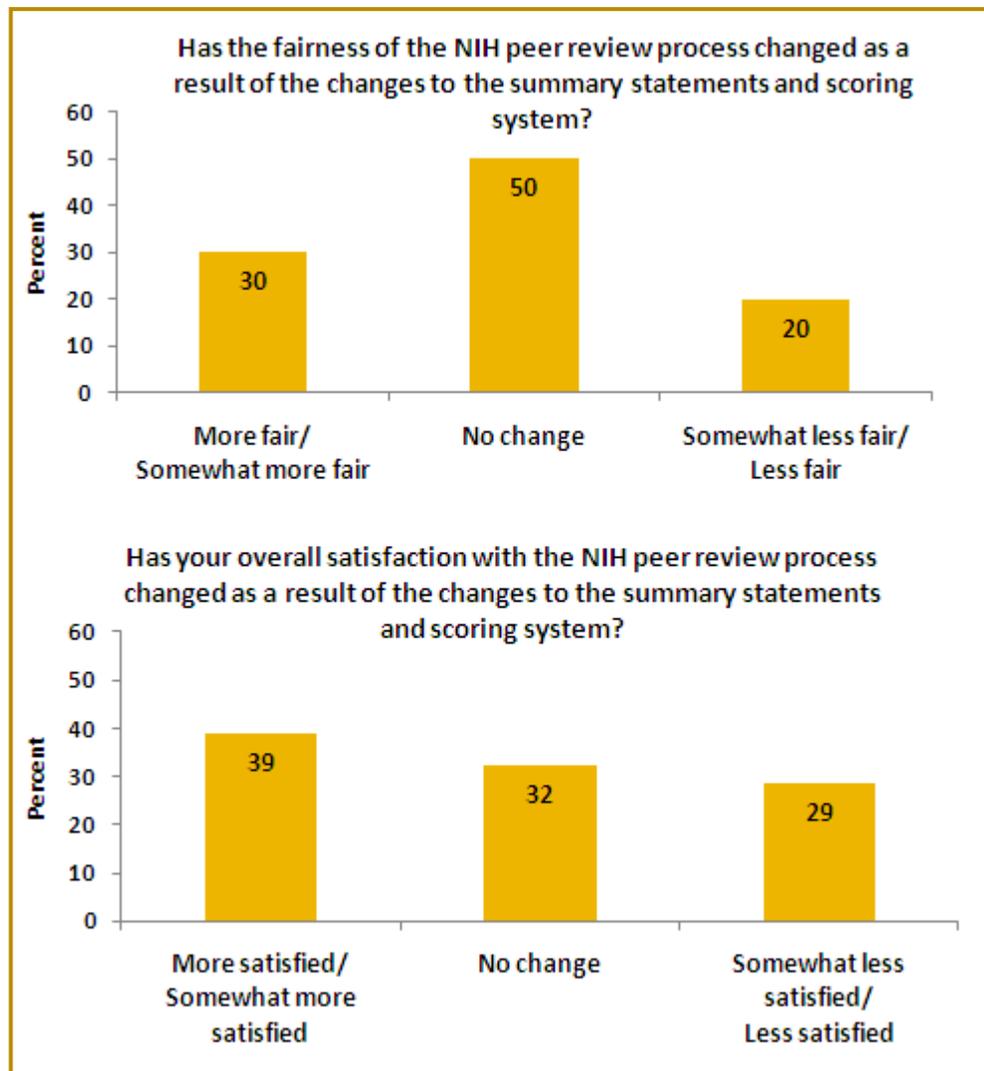
Conclusions

Nine-point Scoring Range: NIH introduced the nine-point scoring range to address concerns that the previous 41-point scoring range depicted a level of precision that was not realistic and that the phenomenon of "score compression" might be an artifact of the inability of reviewers to fully utilize the number of scoring increments available to them.

The Enhancing Peer Review surveys questioned reviewers and Advisory Council members about the adequacy of the nine-point scoring system. The reviewers indicated that the nine-point scoring range was adequate for them to communicate meaningful differences in the quality of applications. Advisory Council members indicated that the nine-point scale was easy for them to understand.

Criterion Scoring: NIH adopted criterion scoring to communicate quantitative ratings of merit from the assigned reviewers for all applications, including those not discussed by the full committee. The assigned reviewers and discussant(s) assign numeric scores on the nine-point scale for each of five review criteria, and these criterion scores are reported individually on the summary statement.

The Enhancing Peer Review surveys examined whether reviewers and applicants found criterion scores helpful to applicants for communicating and understanding the reviewers' assessment of the scientific merit of the applications. Forty-nine percent of reviewers strongly agreed/agreed that criterion scores were helpful to them in communicating why an application was not discussed. Additionally, Program Officers rated criterion scores as one of the Enhancing Peer Review changes that had been most helpful for advising applicants after review. However, applicants' responses reflected no clear agreement about whether criterion scores were helpful for understanding the strengths and weaknesses of the application or the problem areas that could be corrected.

The criterion scores reflect the individual reviewers' ratings of each of the review criteria, whereas the overall impact/priority score is an aggregate score that reflects the ratings assigned by all eligible study section members after discussion of the strengths and weaknesses of the application in study section. The surveys included questions that examined the consistency of the criterion scores with overall impact scores. SROs and POs disagreed/strongly disagreed more often than they agreed that criterion scores were consistent with overall impact/priority scores. These results suggest that the NIH may want to explore ways to improve the helpfulness of criterion scores for all stakeholders.

Bulleted Critiques: NIH instructed reviewers to prepare their critiques as bulleted lists of strengths and weaknesses to focus the review on the factors that influence the merit of the application. NIH standardized the format of critiques by creating fillable critique templates to help reviewers develop their critiques.

The Enhancing Peer Review surveys examined whether the bulleted critique format was effective at communicating the factors that influenced the score of the application. Applicants agreed that both the new and old summary statement formats helped them to focus on problem areas that could be corrected. However, reviewers who rated the peer review system after the Enhancing Peer Review changes were introduced were significantly less likely to strongly agree/agree than reviewers who rated

Conclusions (continued)

the old peer review system that the bulleted critique format was adequate for capturing strengths and weaknesses in comparison to the former, narrative format. SROs' responses reflected no clear agreement about whether the bulleted critique format was useful for focusing the critiques on factors that influence score, and POs disagreed/strongly disagreed more often than they strongly agreed/agreed that summary statements were helpful to them for explaining the recommendations of the review group.

The Enhancing Peer Review surveys also examined whether the new, bulleted critique format effectively communicated the reason(s) applications were not discussed. Applicants, SROs and POs disagreed/strongly disagreed more often than they strongly agreed /agreed that the new summary statement format helped them to understand why applications were not discussed. Reviewers who were asked to rate the new bulleted critique format agreed significantly less often and disagreed significantly more often that the format was helpful for communicating to applicants why their applications were not discussed.

The Enhancing Peer Review surveys asked reviewers to rate whether the critique formats were helpful for completing their critiques efficiently. Reviewers agreed/strongly agreed more often than they disagreed/strongly disagreed that both formats were helpful to them. However, reviewers who rated the new, structured critique templates disagreed/strongly disagreed significantly less often that the format was helpful in completing their critiques efficiently in comparison to reviewers who rated the old, narrative format.

Taken together, these results suggest that the new structured critique templates have benefitted reviewers in terms of efficiency. However, few stakeholders rated the bulleted critique format as helpful for communicating information about the pertinent factors that affected the outcome of the review. These results suggest that the NIH may want to explore ways to improve the helpfulness of bulleted critiques for all stakeholders.

Enhanced Review Criteria: NIH developed the enhanced review criteria to provide clearer guidance to reviewers on the factors that should be considered in the assessment of scientific merit. However, SROs and POs were more likely to strongly disagree than to strongly agree that the enhanced review criteria resulted in greater clarity regarding the strengths and weaknesses of the application.

Clustering of Clinical Research and New Investigator Applications: The practice of clustering was formalized for applications proposing clinical research that were assigned to "mixed" (clinical and non-clinical) study sections and for applications submitted by New Investigators. Most reviewers strongly agreed/agreed that clustering resulted in a more consistent review of affected applications. SROs and POs selected clustering most frequently as a change that had contributed positively to the Enhancing Peer Review objectives. Thus, the surveys indicate that clustering was a positive change brought about by the Enhancing Peer Review Initiative.

Conclusions (continued)

Overall Satisfaction: Applicants who were asked to rate the peer review system after the Enhancing Peer Review changes were introduced expressed no significant preference for the new system over the old. The pattern of results for applicants who rated the peer review system before and after the changes was similar. Applicants rated the peer review system as fair or very fair most often and rated themselves as satisfied or very satisfied.

Most reviewers who were asked to rate the peer review system after the Enhancing Peer Review changes were introduced expressed a preference for the new peer review system over the old system. Although reviewers rated the new peer review system as very fair/fair significantly less often, and rated themselves as very satisfied/satisfied with the new peer review system significantly less often than reviewers asked to rate the old peer review system, reviewers in both groups reported high levels of fairness and satisfaction.

SROs expressed no clear preference for the old peer review system versus the new and most SROs responded that the peer review system remains very fair or fair after the Enhancing Peer Review changes. Approximately equal numbers of SROs rated themselves as very satisfied/satisfied as rated themselves dissatisfied/very dissatisfied.

POs slightly preferred the old peer review system over the new system and their responses to questions about fairness and satisfaction were very similar to those for SROs: most POs responded that the new peer review system was very fair/fair, although approximately equal numbers of POs rated themselves as very satisfied/satisfied as rated themselves dissatisfied/very dissatisfied.

Advisory Council members expressed a moderate preference for the new system over the old. Fifty percent of Advisory Council members indicated that their perceptions of the fairness of the peer review system had not changed since the introduction of the Enhancing Peer Review changes and approximately equal numbers rated their own satisfaction as having improved, remained the same or worsened since the Enhancing Peer Review changes were introduced.

Refinements Already Implemented: NIH has already implemented refinements to the peer review system in response to formal feedback from the Enhancing Peer Review surveys and informal feedback from NIH staff members. These refinements include:

> **May 2009:** Removed electronic barriers in the Internet Assisted Review system to allow reviewers to update criterion scores more easily.
> **July 2009:** Removed the guidance limiting critiques to one-quarter page for each criterion.
> **November 2009:** Modified scoring descriptors by removing a graphic**.**
> **January 2010:** Provided guidance clarifying Overall Impact versus significance.
> **September 2010:** Required reviewers to include a narrative statement to explain the Overall Impact score.
> **September 2010:** Added footnote to summary statement explaining that the criterion scores were entered before the study section meeting.

Conclusions (continued)

Continuous Review of Peer Review: The Enhancing Peer Review Initiative was conceived as an ongoing process of refinement to ensure that the NIH peer review system continues to evolve as science evolves. Continual monitoring and assessment are needed to facilitate this process. The Enhancing Peer Review surveys presented here were an early snapshot of the opinions of NIH stakeholders about the changes to the peer review process that were implemented in May 2009. Only one complete application cycle had occurred when the applicant and reviewer surveys were deployed in December 2009, and NIH staff members had administered two full application cycles prior to being surveyed in April 2010. NIH will continue to monitor stakeholder opinions of the scoring system, the critique format and other peer review processes at a later date. In addition, NIH's ongoing review of the NIH peer review system will also examine the shortened applications and the alignment of the application with the NIH review criteria.

Enhancing Peer Review Refinements

Allowed scores to be revised without re-uploading critiques

Appl & Rev Surveys

Council Surveys

Converted format of Overall Impact critique from bullets to paragraph

Modified scoring guidance (removed graphic)

SRO & PO Surveys

Provided guidance on Overall Impact v. Significance

| 2009 | May-Jun | Jul-Aug | Sep-Oct | Nov-Dec | Jan-Feb | Mar-Apr | May-Jun | Jul-Aug | Sep-Oct | 2010 |

Removed guidance to limit critiques to one-quarter page

Implementation of Administrative Note for noncompliant applications

Added an explanation to summary statements about criterion scores

*Denotes refinements that were enacted as a result of feedback from the surveys.

http://enhancing-peer-review.nih.gov/

APPENDIX I – How the Surveys Were Conducted

The Enhancing Peer Review surveys focused only on the Stage 2 changes to the peer review system and were launched in January of 2010.

Applicant and Reviewer Surveys (December - January 2010):
- Web-based instruments
- OMB-approved customer satisfaction surveys
- Sampled two groups of applicants and reviewers:
  1. Applicants and/or reviewers who had participated in the peer review system BEFORE the changes were introduced (Number of respondents: 374 Applicants and 221 Reviewers)
  2. Applicants and/or reviewers who had participated in the peer review system AFTER the changes were introduced (Number of respondents: 504 Applicants and 537 Reviewers)
- Participation in ARRA FOAs and review meetings was not counted in drawing the sampling frame, but these individuals also were not excluded. Respondents were asked about their participation in the ARRA reviews.

Analysis of Applicant – Reviewer Survey questions:
Initial analysis compared responses about reviews that took place BEFORE (not experienced) vs. AFTER the changes (experienced).

- **Secondary multinomial analysis examined demographic influences on survey questions**
  - Experienced vs. not experienced
  - Whether applicants' most recent application was funded
  - Whether application proposed clinical research
  - Whether reviewers have experience as applicants
  - Gender
  - Age (45 and younger; 46 and older)
  - Professional rank (full professor; associate professor/senior scientist; or other)
  - Institution type (university versus other)
  - Education (Ph.D. versus other)
  - Year first NIH grant application was submitted (1990 or earlier; 1991-present)

Note: Race and ethnicity were initially examined but the sample sizes for underrepresented minorities were too small for a valid analysis to be conducted, so this factor was dropped from the analysis.

Advisory Council Surveys (January - February 2010):
- Paper and pencil instrument administered during the February Advisory Council/Board meetings
- Were treated as employee satisfaction surveys (no OMB approval)
- 291 respondents, all serving in the current meeting and thus experienced AFTER the changes were introduced

APPENDIX I – How the Surveys Were Conducted (continued)

SROs and POs (April - May 2010)
- Web-based instruments
- Were treated as employee satisfaction surveys (no OMB approval)
- 288 SRO respondents and 437 PO respondents
- All eligible SROs and POs had administered applications AFTER the changes were introduced

Analysis of SRO, PO and Advisory Council surveys is descriptive only.

The current report focuses only on the changes to the peer review system. Other questions examined:

- Demographic factors (Applicant, Reviewer)
- Funding history and review service history (Applicant, Reviewer)
- Factors affecting willingness to review (Reviewer)
- Factors affecting effort spent on preparing a grant application (Applicant)
- Prior employment experience at NIH (SRO/PO)
- Workload factors such as type and number of applications administered
- Factors affecting effort spent recruiting reviewers (SROs)
- Factors affecting effort spent advising applicants (POs)
- Usability of training materials and automated resources provided to support the Enhancing Peer Review Initiative

APPENDIX II – Current Guidelines for Reviewers Including Scoring Descriptors

This scoring system was designed to encourage more reliable scoring of applications. Highly rating all applications greatly diminishes the ability of a reviewer or study section to communicate the scientific impact of an application. Therefore, reviewers who carefully consider the rating guidance provided in determining their scores improve not only the reliability of their scores, but also improve their ability to communicate the scientific impact of the applications reviewed.

Contents

Scoring

Scoring

Summary

- The NIH grant application scoring system uses a nine-point scale
- A score of 1 indicates an exceptionally strong application with essentially no weaknesses. A score of 9 indicates an application with serious and substantive weaknesses with very few strengths; 5 is considered an average score
- Ratings are in whole numbers only (no decimal ratings)
- This scale is used by all eligible (without conflict of interest) Scientific Review Group members to provide an overall impact/priority score and for assigned reviewers to score five individual criteria (e.g., Significance, Investigator(s), Innovation, Approach, Environment)
- For the impact/priority score rating, strengths and weaknesses across all of the review criteria should be considered
    - For each criterion rating, the strengths and weaknesses within that review criterion should be considered
- Reviewers should consider not only the relative number of strengths and weaknesses noted, but also the importance of these strengths and weaknesses to the criteria or to the overall impact when determining a score
    - For example, a major strength may outweigh many minor and correctable weaknesses
- For information about using the critique template, see Critique Template Instructions
- NIH expects that scores of 1 or 9 would be used less frequently than the other scores

http://enhancing-peer-review.nih.gov/

APPENDIX II – Current Guidelines for Reviewers Including Scoring Descriptors (continued)

Preliminary Scores
- Before the review meeting, assigned reviewers will determine preliminary scores for each of the five scored review criteria and a preliminary score for the overall impact/priority
- The impact/priority score should reflect the reviewer's overall evaluation, not a numerical average of individual criterion scores
- Reviewers should consider the full range of the rating scale and the scoring descriptors in assigning preliminary and final scores
  - However, a reviewer should not assume that the applications assigned to him/her necessarily cover that entire range of scores, and should assign scores as appropriate for the work or science proposed
- An application does not need to be strong in all categories to be judged likely to have major impact
  - For example, a project that by its nature is not innovative may be essential to advance a field
- Reviewers must enter the criterion scores into the Internet Assisted Review (IAR) site in the NIH Commons for them to appear in the summary statement
  - If entered in IAR, the scores will be transferred to a table at the beginning of the reviewer's critique
- Assigned reviewers may submit criterion scores only after their critiques have been uploaded
  - At the SRO's discretion, discussants who are assigned to the application and SRG members who are not assigned to the application may submit criterion scores without critiques
- In the READ phase of the meeting reviewers may submit their scores and critiques, but may not edit them
- These preliminary scores are not retained, but will be replaced by final scores that are given by private scoring and are based on the outcome of the deliberations at the peer review meeting

Criterion Scoring
- In most cases, up to five individual criteria are scored, but certain funding opportunity announcements may include more than five scored criteria
- Criterion scores are provided for both discussed and not discussed applications
- Criterion scores are intended to provide additional information on how each assigned reviewer weighed that particular section so that the reader has a better idea of strengths and weaknesses that need improvement
- Providing scores without providing comments in the review critique is discouraged
- The impact/priority score for the application is not intended to be an average of criterion scores
- Criterion scores are entered into the Internet Assisted Review site for the meeting; the same screen also allows uploading of the written critique at the same time
- If the reviewer's opinion changed as a result of discussion at the meeting, the reviewer should change his/her criterion scores to match his/her critiques and overall impact/priority score
- The criterion scores appear in a table at the beginning of each critique in the summary statement

APPENDIX II – Current Guidelines for Reviewers Including Scoring Descriptors (continued)

Impact/Priority Score
- Discussed applications will receive numerical impact/priority scores from all eligible reviewers (e.g., without conflicts of interest)
- The impact/priority score for an application is based on each individual reviewer's assessment based on the five scored criteria plus additional criteria regarding the protection and inclusion of human subjects, vertebrate animal care and welfare, biohazards, and criteria specific to the application
- Reviewers are guided to use the full range of the rating scale and spread their scores to better discriminate among applications
- Reviewers whose evaluations or opinions of an application fall outside the range of those presented by the assigned reviewers and discussant(s) should ensure that their opinions are brought to the attention of the entire committee
- In addition, the SRO and chairperson should ensure that all opinions are voiced before final scoring is conducted
- Reviewers should feel free to assign the score that they believe best represents the impact of the application, and not feel constrained to limit their scores to the upper half of the score range if they do not feel such a score is warranted
- After the meeting, individual reviewer scores will be averaged and the result multiplied by 10 to determine the final impact/priority score
- The range of the final application scores is from 10 to 90

Non-numeric Scores
- Not Discussed (ND)
    - Applications unanimously judged by the peer review committee to be less competitive are not discussed at the peer review meeting
    - These applications do not receive a numerical impact/priority score
    - These applications do receive individual criterion scores
    - No set number of applications are discussed; in some meetings, the "Not Discussed" option may not be used
- Not Recommended for Further Consideration (NRFC)
    - NRFC for an application occurs by majority vote of the peer reviewers
    - NRFC occurs in the following scenarios:
        - Application lacks significant and substantial merit
        - Application presents serious ethical problems in the protection of human subjects from research risks
        - Application presents serious ethical problems in the use of vertebrate animals, biohazards and/or select agents
    - NRFC-scored applications do not proceed to the second level of peer review (National Advisory Council/Board) because they cannot be funded
    - The NRFC is a serious committee recommendation that is substantially different from Not Discussed (ND)

APPENDIX II – Current Guidelines for Reviewers Including Scoring Descriptors (continued)

- Other Non-numeric Scores
  - Deferred (usually due to lack of sufficient information, quorum, allegations of research misconduct)
  - Abstention (used rarely)
  - Conflict (score put in by a reviewer who is in conflict with the application)
  - Not present

Reviewer Guidance and Chart
- For the impact/priority score and for the individual criterion scores, the far right column (in the table below) provides a descriptive guide of how strengths and weaknesses are considered in assigning a rating
  - **Minor weakness**: easily addressable weakness, does not substantially lessen impact
  - **Moderate weakness**: lessens impact
  - **Major weakness**: Severely limits impact
- Impact (far left column) is the project's likelihood to have a sustained, powerful influence on the research field(s) involved
  - High impact = 1 to 3
  - Moderate impact = 4 to 6
  - Low impact = 7 to 9
- Each review criterion should be assessed based on how important each review criterion is to the work being proposed
  - As a result, a reviewer may give only moderate scores to some of the review criteria but still give a high overall impact/priority score because the one review criterion critically important to the research is rated highly; or a reviewer could give mostly high criterion ratings but rate the overall impact/priority score lower because the one criterion critically important to the research being proposed is not highly rated.
- An application does not need to be strong in all categories to be judged likely to have major impact; e.g., a project that by its nature is not innovative may be essential to advance a field.

Impact	Score	Descriptor	Additional Guidance on Strengths/Weaknesses
High	1	Exceptional	Exceptionally strong with essentially no weaknesses
	2	Outstanding	Extremely strong with negligible weaknesses
	3	Excellent	Very strong with only some minor weaknesses
Medium	4	Very Good	Strong but with numerous minor weaknesses
	5	Good	Strong but with at least one moderate weakness
	6	Satisfactory	Some strengths but also some moderate weaknesses
Low	7	Fair	Some strengths but with at least one major weakness
	8	Marginal	A few strengths and a few major weaknesses
	9	Poor	Very few strengths and numerous major weaknesses

APPENDIX II – Current Guidelines for Reviewers Including Scoring Descriptors (continued)

Additional Information for Scoring Guidance Table
Non-numeric score options: NR = Not Recommended for Further Consideration, DF = Deferred, AB = Abstention, CF = Conflict, NP = Not Present, ND = Not Discussed
Minor weakness: An easily addressable weakness that does not substantially lessen impact **Moderate weakness:** A weakness that lessens impact **Major weakness:** A weakness that severely limits impact

Percentiling

- For the appropriate applications (certain activity codes or RFAs), scores will be percentiled to the appropriate base (e.g. study section base if the number of R01 applications $\geq$ 25; CSR-all or IC-all base if <25)
- All percentiles are rounded to a whole number
- Until a base has been established from three rounds of review (i.e., May 2010 Council), percentiles are based on less than three application rounds

APPENDIX III – Side-by-Side Comparison of Enhanced and Former Review Criteria

Section	Former Review Criteria (NOT-OD-05-002 and NOT-OD-06-069)	Enhanced Review Criteria (NOT-OD-09-025)
Introduction	The goals of NIH supported research are to advance our understanding of biological systems, to improve the control of disease, and to enhance health. In their written critiques, reviewers will be asked to comment on each of the following criteria in order to judge the likelihood that the proposed research will have a substantial impact on the pursuit of these goals. Each of these criteria will be addressed and considered in assigning the overall score, weighting them as appropriate for each application. Note that an application does not need to be strong in all categories to be judged likely to have major scientific impact and thus deserve a high priority score. For example, an investigator may propose to carry out important work that by its nature is not innovative but is essential to move a field forward.	The mission of the NIH is to support science in pursuit of knowledge about the biology and behavior of living systems and to apply that knowledge to extend healthy life and reduce the burdens of illness and disability. As part of this mission, applications submitted to the NIH for grants or cooperative agreements to support biomedical and behavioral research are evaluated for scientific and technical merit through the NIH peer review system. **Overall Impact.** Reviewers will provide an overall impact/priority score to reflect their assessment of the likelihood for the project to exert a sustained, powerful influence on the research field(s) involved, in consideration of the following five core review criteria, and additional review criteria (as applicable for the project proposed).
Scored Review Criteria	**Significance:** Does this study address an important problem? If the aims of the application are achieved, how will scientific knowledge or clinical practice be advanced? What will be the effect of these studies on the concepts, methods, technologies, treatments, services, or preventative interventions that drive this field? **Approach:** Are the conceptual or clinical framework, design, methods, and analyses adequately developed, well-integrated, well-reasoned, and appropriate to the aims of the project? Does the applicant acknowledge potential problem areas and consider alternative tactics? For applications designating multiple PIs, does the Leadership Plan ensure that there will be sufficient coordination and communication among the PIs? Are the administrative plans for the management of the research project appropriate, including plans for resolving conflicts? **Innovation:** Is the project original and innovative? For example: Does the project challenge existing paradigms or clinical practice; address an innovative hypothesis or critical barrier to progress in the field? Does the project develop or employ novel concepts, approaches or methodologies, tools, or technologies for this area? **Investigators:** Are the principal investigator(s) and key	**Scored Review Criteria.** Reviewers will consider each of the five review criteria below in the determination of scientific and technical merit, and give a separate score for each. An application does not need to be strong in all categories to be judged likely to have major scientific impact. For example, a project that by its nature is not innovative may be essential to advance a field. *Significance.* Does the project address an important problem or a critical barrier to progress in the field? If the aims of the project are achieved, how will scientific knowledge, technical capability, and/or clinical practice be improved? How will successful completion of the aims change the concepts, methods, technologies, treatments, services, or preventative interventions that drive this field? *Investigator(s).* Are the PD/PIs, collaborators, and other researchers well suited to the project? If Early Stage Investigators or New Investigators, do they have appropriate experience and training? If established, have they demonstrated an ongoing record of accomplishments that have advanced their field(s)? If the project is collaborative or multi-PD/PI, do the investigators have complementary and integrated expertise; are their leadership approach, governance and organizational structure appropriate for the project? *Innovation.* Does the application challenge and seek to shift current research or clinical practice paradigms by utilizing novel theoretical concepts, approaches or methodologies, instrumentation, or interventions? Are the concepts, approaches or methodologies, instrumentation, or interventions novel to one field of research or novel in a broad sense? Is a refinement, improvement, or new application of theoretical concepts, approaches or methodologies, instrumentation, or interventions proposed? *Approach.* Are the overall strategy, methodology, and analyses well-reasoned and appropriate to accomplish the specific aims of the project? Are potential problems, alternative strategies, and benchmarks for success presented? If the project is in the early stages of development, will the

APPENDIX III – Side-by-Side Comparison of Enhanced and Former Review Criteria (continued)

Section	Former Review Criteria (NOT-OD-05-002 and NOT-OD-06-069)	Enhanced Review Criteria (NOT-OD-09-025)
	personnel appropriately trained and well suited to carry out this work? Is the work proposed appropriate to the experience level(s) of the principal investigator(s) and other researchers? Do the principal investigator(s) and investigative team bring complementary and integrated expertise to the project (if applicable)? **Environment:** Does the scientific environment in which the work will be done contribute to the probability of success? Do the proposed studies benefit from unique features of the scientific environment(s), or subject populations, or employ useful collaborative arrangements? Is there evidence of institutional support?	strategy establish feasibility and will particularly risky aspects be managed? If the project involves clinical research, are the plans for 1) protection of human subjects from research risks, and 2) inclusion of minorities and members of both sexes/genders, as well as the inclusion of children, justified in terms of the scientific goals and research strategy proposed? *Environment.* Will the scientific environment in which the work will be done contribute to the probability of success? Are the institutional support, equipment and other physical resources available to the investigators adequate for the project proposed? Will the project benefit from unique features of the scientific environment, subject populations, or collaborative arrangements?
Additional Review Criteria	**Protection of Human Subjects from Research Risk:** The involvement of human subjects and protections from research risk relating to their participation in the proposed research will be assessed (see the Research Plan, Section E on Human Subjects in the PHS Form 398). **Inclusion of Women, Minorities and Children in Research:** The adequacy of plans to include subjects from both genders, all racial and ethnic groups (and subgroups), and children as appropriate for the scientific goals of the research will be assessed. Plans for the recruitment and retention of subjects will also be evaluated (see the Research Plan, Section E on Human Subjects in the PHS Form 398). **Care and Use of Vertebrate Animals in Research:** If vertebrate animals are to be used in the project, the five items described under Section F of the PHS Form 398 research grant application instructions will be assessed	**Additional Review Criteria.** As applicable for the project proposed, reviewers will consider the following additional items in the determination of scientific and technical merit, but will not give separate scores for these items. *Protections for Human Subjects.* For research that involves human subjects but does not involve one of the six categories of research that are exempt under 45 CFR Part 46, the committee will evaluate the justification for involvement of human subjects and the proposed protections from research risk relating to their participation according to the following five review criteria: 1) risk to subjects, 2) adequacy of protection against risks, 3) potential benefits to the subjects and others, 4) importance of the knowledge to be gained, and 5) data and safety monitoring for clinical trials. For research that involves human subjects and meets the criteria for one or more of the six categories of research that are exempt under 45 CFR Part 46, the committee will evaluate: 1) the justification for the exemption, 2) human subjects involvement and characteristics, and 3) sources of materials. *Inclusion of Women, Minorities, and Children.* When the proposed project involves clinical research, the committee will evaluate the proposed plans for inclusion of minorities and members of both genders, as well as the inclusion of children. *Vertebrate Animals.* The committee will evaluate the involvement of live vertebrate animals as part of the scientific assessment according to the following five points: 1) proposed use of the animals, and species, strains, ages, sex, and numbers to be used; 2) justifications for the use of animals and for the appropriateness of the species and numbers proposed; 3) adequacy of veterinary care; 4) procedures

APPENDIX III – Side-by-Side Comparison of Enhanced and Former Review Criteria (continued)

Section	Former Review Criteria (NOT-OD-05-002 and NOT-OD-06-069)	Enhanced Review Criteria (NOT-OD-09-025)
		for limiting discomfort, distress, pain and injury to that which is unavoidable in the conduct of scientifically sound research including the use of analgesic, anesthetic, and tranquilizing drugs and/or comfortable restraining devices; and 5) methods of euthanasia and reason for selection if not consistent with the AVMA Guidelines on Euthanasia. *Biohazards.* Reviewers will assess whether materials or procedures proposed are potentially hazardous to research personnel and/or the environment, and if needed, determine whether adequate protection is proposed. *Resubmission Applications.* When reviewing a Resubmission application (formerly called an amended application), the committee will evaluate the application as now presented, taking into consideration the responses to comments from the previous scientific review group and changes made to the project. *Renewal Applications.* When reviewing a Renewal application (formerly called a competing continuation application), the committee will consider the progress made in the last funding period. *Revision Applications.* When reviewing a Revision application (formerly called a competing supplement application), the committee will consider the appropriateness of the proposed expansion of the scope of the project. If the Revision application relates to a specific line of investigation presented in the original application that was not recommended for approval by the committee, then the committee will consider whether the responses to comments from the previous scientific review group are adequate and whether substantial changes are clearly evident.

APPENDIX III – Side-by-Side Comparison of Enhanced and Former Review Criteria (continued)

Section	Former Review Criteria (NOT-OD-05-002 and NOT-OD-06-069)	Enhanced Review Criteria (NOT-OD-09-025)
Additional Review Considerations	**Budget:** The reasonableness of the proposed budget and the requested period of support in relation to the proposed research. The priority score should not be affected by the evaluation of the budget.	**Additional Review Considerations.** As applicable for the project proposed, reviewers will address each of the following items, but will not give scores for these items and should not consider them in providing an overall impact/priority score. *Budget and Period Support.* Reviewers will consider whether the budget and the requested period of support are fully justified and reasonable in relation to the proposed research. *Select Agent Research.* Reviewers will assess the information provided in this section of the application, including 1) the Select Agent(s) to be used in the proposed research, 2) the registration status of all entities where Select Agent(s) will be used, 3) the procedures that will be used to monitor possession use and transfer of Select Agent(s), and 4) plans for appropriate biosafety, biocontainment, and security of the Select Agent(s). *Applications from Foreign Organizations.* Reviewers will assess whether the project presents special opportunities for furthering research programs through the use of unusual talent, resources, populations, or environmental conditions that exist in other countries and either are not readily available in the United States or augment existing U.S. resources. *Resource Sharing Plans.* Reviewers will comment on whether the following Resource Sharing Plans, or the rationale for not sharing the following types of resources, are reasonable: 1) Data Sharing Plan (http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm);2) Sharing Model Organisms (http://grants.nih.gov/grants/guide/notice-files/NOT-OD-04-042.html); and 3) Genome Wide Association Studies (GWAS) (http://grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html).

APPENDIX III – Side-by-Side Comparison of Enhanced and Former Review Criteria (continued)

	Research and Research Center (R, DP, RC, P, etc)	SBIR/STTR (R41, R42, R43, R44)	Fellowship (F30, F31, F32, F33)	Career Development (K01, K02, K07, K08, K23, K24, K25, K99)	Institutional Training (T32)	Shared Instrumentation (S10)
Overall Impact	*Overall Impact*	*Overall Impact*	*Overall Impact/Merit*	*Overall Impact*	*Overall Impact*	*Overall Impact/Benefit*
Scored Review Criteria (Scored individually and considered in overall impact/priority score)	✓ *Significance* ✓ *Investigator(s)* ✓ *Innovation* ✓ *Approach* ✓ *Environment* PAR & RFA: May add questions to each scored or additional criterion — FOA-specific — Not given individual criterion scores	✓ *Significance* ✓ *Investigator(s)* ✓ *Innovation* ✓ *Approach* ✓ *Environment*	✓ *Fellowship Applicant* ✓ *Sponsors, Collaborators, and Consultants* *Research Training Plan* *Training Potential* *Institutional Environment & Commitment to Training*	✓ *Candidate* ✓ *Career Development Plan/Career Goals & Objectives/Plan to Provide Mentoring* ✓ *Research Plan* *Mentor(s), Co-Mentor(s), Consultant(s), Collaborator(s)* ✓ *Environment & Institutional Commitment to the Candidate*	✓ *Training Program and Environment* ✓ *Training PD/PI* ✓ *Preceptors /Mentors* ✓ *Trainees* ✓ *Training Record* Other T programs use other criteria	✓ *Justification of Need* ✓ *Technical Expertise* ✓ *Research Projects* ✓ *Administration* ✓ *Institutional Commitment* • *Overall Benefit (not scored)*
Additional Review Criteria (Not scored individually, but considered in overall impact/priority score)	R01-BRP only: • *Partnership and Leadership* All: ✓ *Protections for Human Subjects* ✓ *Inclusion of Women, Minorities, & Children* ✓ *Vertebrate Animals* ✓ *Biohazards* • *Resubmission* • *Renewal* • *Revision*	• *Phase II* • *Fast Track* ✓ *Protections for Human Subjects* ✓ *Inclusion of Women, Minorities, & Children* ✓ *Vertebrate Animals* ✓ *Biohazards* • *Resubmission* • *Renewal* • *Revision*	✓ *Protections for Human Subjects* ✓ *Inclusion of Women, Minorities, & Children* ✓ *Vertebrate Animals* ✓ *Biohazards* • *Resubmission* • *Renewal*	✓ *Protections for Human Subjects* ✓ *Inclusion of Women, Minorities, & Children* ✓ *Vertebrate Animals* ✓ *Biohazards* • *Resubmission* • *Renewal* • *Revision*	✓ *Protections for Human Subjects* ✓ *Inclusion of Women, Minorities, & Children* ✓ *Vertebrate Animals* ✓ *Biohazards* • *Resubmission* • *Renewal* • *Revision*	✓ *Biohazards* ✓ *Resubmission*
Additional Review Considerations (Not scored individually and not considered in overall score)	R01-BRP only: ✓ *Technology Transfer* All: • *Applications from Foreign Organizations* • *Select Agents* • *Resource Sharing Plans* ✓ *Budget & Period of Support*	• *Select Agents* • *Resource Sharing Plans* ✓ *Budget & Period of Support*	✓ *Training in the Responsible Conduct of Research* • *Applications from Foreign Organizations* • *Select Agents* • *Resource Sharing Plans* ✓ *Budget & Period of Support*	✓ *Training in the Responsible Conduct of Research* • *Select Agents* • *Resource Sharing Plans* ✓ *Budget & Period of Support*	✓ *Recruitment & Retention Plan to Enhance Diversity* ✓ *Training in the Responsible Conduct of Research* • *Select Agents* ✓ *Budget & Period of Support*	✓ *Budget & Period of Support*
Additional Comments to Applicant	Additional Comments to Applicant	Additional Comments to Applicant	Additional Comments to Applicant	Additional Comments to Applicant	Additional Comments to Applicant	Additional Comments to Applicant

http://enhancing-peer-review.nih.gov/