Scoring System and Procedure

REVIEWER TRAINING SUMMARY PAGE

- The NIH grant application scoring system is being implemented to improve rating reliability, encourage use of the full scoring range, and provide quantitative feedback on all applications, both discussed and not discussed.
- The NIH grant application scoring system uses a 9-point rating for the impact/priority score with 1 = Exceptional and 9 = Poor.
- Ratings are in whole numbers only (no decimal ratings).
- Assigned reviewers also provide ratings for each review criterion [e.g. Significance, Investigator(s), Innovation, Approach, Environment] using the same 9-point scale.
    - These criterion ratings are provided in the summary statement for applications, both discussed and not discussed.
    - Criterion ratings should be considered in determining the overall impact/priority score, but reviewers should determine the relative importance of each criterion for the science or work being proposed.
- Reviewers should use the full range of the rating scale and spread their scores to better discriminate among applications.
- Discussed applications will receive impact/priority scores from all eligible reviewers (e.g., without conflicts of interest). Individual reviewer scores will be averaged and the result multiplied by 10 to determine the final impact/priority score (range of 10 to 90).
- Scores will be percentiled to the appropriate base (e.g. study section base if the number of applications $\geq$25; CSR all base, or IC all base if < 25) and reported in whole number percentiles. Until a new base has been established from three rounds of reviews, percentiles will be based only on the current round of applications (reviews for October 2009 Council) or the prior and current rounds (reviews for January 2010 Council).

DETAILED INSTRUCTIONS

Rationale for the New NIH Grant Application Scoring System

The prior scoring system of 1.0 to 5.0 in 0.1 increments served NIH well for many years, but its weaknesses became increasingly evident as the quality and quantity of applications increased and NIH budgets to fund grant applications tightened. The new scoring system is being implemented to address the following issues:

- For even the most experienced reviewers, it is difficult to make 41 reliable discriminations of application merit. Based on measurement science, prior experience, and feedback from various constituencies, a 9-point rating scale with descriptors associated with each rating option was adopted.
- Reviewer ratings became increasingly positive, compressing the score range, and effectively reducing the usefulness of scores for NIH funding decisions. In the new scoring system, the descriptors associated with each rating were designed to encourage use of the full scoring range.
- To provide additional feedback to applicants, program staff, and other consumers of the summary statement, assigned reviewers also provide rating of the specific review criteria using the same 9-point scale.

The NIH Grant Application Scoring System

The NIH scoring system uses a 9-point rating scale from 1 = Exceptional to 9 = Poor for the overall impact/priority score as well as the individual review criteria. Ratings are provided only in whole numbers, not decimals. In addition to the descriptors associated with each rating, two additional rating guides (see below) are provided:

- For the impact/priority score, the far left column provides guidance for assigning scores to applications based on the project's likelihood to have a sustained, powerful influence on the research field(s) involved:

  1 to 3 = high impact
  4 to 6 = moderate impact
  7 to 9 = low impact

- For the impact/priority score and for the individual criterion scores, the far right column provides a graphical guide of how strengths and weaknesses are considered in assigning a rating. A score of 1 indicates an exceptionally strong application (or exceptionally strong significance, investigators, innovation, approach, environment) with essentially no weaknesses. A score of 9 indicates serious and substantive weaknesses with very few strengths. 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. In considering strengths and weaknesses, reviewers should consider the relative importance of the strengths and weaknesses noted, not simply the number of strengths and weaknesses.

9-Point Score Chart

Impact	Score	Descriptor	Strengths/Weaknesses
High Impact	1	Exceptional	Strengths
	2	Outstanding	
	3	Excellent	
Moderate Impact	4	Very Good	
	5	Good	
	6	Satisfactory	
Low Impact	7	Fair	
	8	Marginal	
	9	Poor	Weaknesses

Non-numeric score options: NR = Not Recommended for Further Consideration, DF = Deferred, AB = Abstention, CF = Conflict, NP = Not Present, ND=Not Discussed

Additional Guidance on Strengths and Weaknesses

The graphical representation of strengths and weaknesses (the far right column) is provided to illustrate the relative balance of strengths and weaknesses associated with each rating score. 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. The table below provides additional guidance to assist reviewers in determining their ratings.

Score	Descriptor	Additional Guidance on Strengths / Weaknesses
1	Exceptional	Exceptionally strong with essentially no weaknesses
2	Outstanding	Extremely strong with negligible weaknesses
3	Excellent	Very strong with only some minor weaknesses
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
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

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

Distribution of Scores

With 9 possible rating discriminations, it is imperative that reviewers distribute or spread their scores as widely as possible among applications. The descriptors associated with each rating were designed to encourage the spreading of scores. Therefore, although score distributions may vary by study section, reviewers should use the full range of 1 to 9; the expectation, however, is that there will be few 1s and few 9s.

Because the new scoring system was designed to encourage greater spreading of scores, it is not appropriate to simply convert scores from the old rating scale to the new rating scale. For example, a rating of 2.0 in the former scoring system does not have the same meaning as a 3 in the new scoring system. A rating of 3 in the new scoring system indicates an excellent application of high impact that is very strong with only some minor weaknesses, considerably better than what is typically indicated by a 2.0 rating in the former scoring system.

Highly rating all applications greatly diminishes the ability of a reviewer or study section to communicate the 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 impact of the applications reviewed.

Scoring and Not Discussed Applications

Most study sections discuss only a percentage (usually 50%) of applications assigned to the study section. Typically, these applications have preliminary scores in the better half of the scoring range. Following discussion, however, 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 score to the upper half of the score range if they do not feel such a score is justified. For example, if the assigned reviewers initially score an application as 4, 5, and 6, and subsequent discussion reveals a serious weakness that will substantially lessen the project's impact, then it is appropriate for reviewers to give a higher (worse) score.

Scoring Range

After discussion, the assigned reviewers state their final scores, defining the score range. Based on the discussion, all eligible reviewers also score the application. If reviewers wish to score outside the score range of the assigned reviewers, they should declare that they intend to score outside the range and briefly describe the reason. Any score outside the range of the assigned reviewers should be declared, even if the range is a single score (i.e. all assigned reviewers give the same final score). It is important that all points of view and opinions of reviewers are discussed; therefore, reviewers should feel free to score outside the range based on their determination of the overall impact of the application.

Additional Guidance on Criterion Scoring

Assigned reviewers provide both preliminary impact/priority scores and criterion scores (ratings of each review criteria). These criterion scores are included in the summary statement to give applicants of both discussed and not discussed (i.e. streamlined) applications a sense of how consideration of the review criteria influenced the overall evaluation of the application. However, because the relative importance of each individual criterion to the overall score differs for each application, reviewers should not use a formula of weighted or unweighted averages across applications to determine the overall impact/priority score. In addition, unrated criteria such as human subjects, vertebrate animal care, and RFA-specific criteria also should be considered in determining the overall

impact/priority score. Therefore, each review criterion should be weighed differently for each application depending 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.

Final Impact/Priority Scores and Percentile Scores

Discussed applications will receive impact/priority scores from all eligible reviewers. Individual reviewer scores will be averaged and the result multiplied by 10 to determine the final impact/priority score (range of 10 to 90) reported in the summary statement.

Scores will be percentiled to the appropriate base (e.g. study section base if the number of applications $\geq$25; CSR all or IC all base if < 25) and reported in whole number percentiles. Until a new base has been established from three rounds of reviews, percentiles will be based only on the current round of applications (reviews for October 2009 Council) or the prior and current rounds (reviews for January 2010 Council).