Evaluation Summary of New 1-9 Scoring System

October 13, 2009

Reviewers

Introduction

This first pilot of the 1-9 scoring system was in two small review meetings each related to a Request for Application. Therefore, these evaluation results are very preliminary. We will be doing additional evaluation of this new scoring scale in the future.

Evaluation Participation

Invitations to take an evaluation survey went to the 33 Reviewers who participated in the 2 meetings that piloted the 1-9 scoring. Of these 33 reviewers, 22 completed the questionnaire for a 67% participation rate.

Key Findings

- Virtually every question directly pertaining to the 1-9 scoring system revealed positive opinions from reviewers who participated in the two scoring pilot meetings.
- Reviewers felt the 1-9 scoring system should be useful to applicants for interpreting the written comments they receive.
- The new 1-9 scoring system was not deemed to be an additional burden when compared to the old 1-5 scoring system.
- 100% of respondents indicated they were either "Very Satisfied" (68%) or "Somewhat satisfied" (32%) with the 1-9 scoring system.
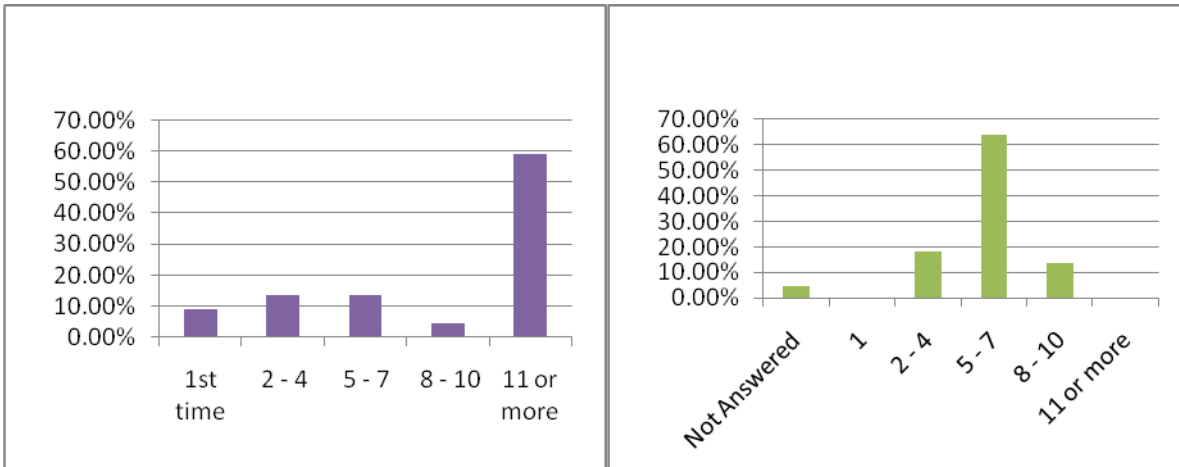
Results

Reviewer Information

Reviewers were asked how many times they had been a reviewer of grant applications for the NIH and how many applications were assigned to them as a reviewer. The majority of respondents had been a reviewer 11 or more times. Over 60% indicated they were assigned 5-7 applications.

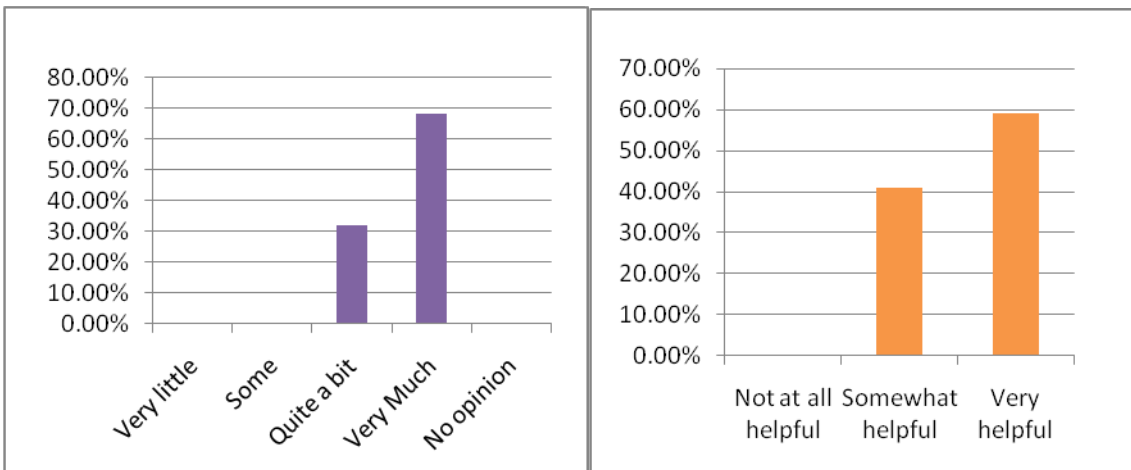How many times have you been a reviewer of grant applications for NIH?

How many applications were assigned to you as a reviewer?

70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%

1st time   2 - 4   5 - 7   8 - 10   11 or more

70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%

Not Answered   1   2 - 4   5 - 7   8 - 10   11 or more

The 1-9 scoring system was found to be effective in helping reviewers to communicate the differences in the impact or merit of the applications, with 100% of the respondents stating it helped "very much" or "quite a bit." The majority of respondents (59%) felt that the "verbal descriptors" were "very helpful" and the remaining repondents indicated the descriptors were at least "somewhat helpful."

How well were you able to communicate the differences in impact or merit of the applications you reviewed using the new 1-9 scoring system?

How helpful were the verbal descriptors (Exceptional to Poor) in your determination of the initial and final scores?

80.00%
70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%

Very little   Some   Quite a bit   Very Much   No opinion

70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%

Not at all helpful   Somewhat helpful   Very helpful

When reviewers were asked to provide recommendations regarding improving the verbal descriptors, nine reviewers responded. Three responses stated that no change should be made to the descriptors. Other comments suggested more clarity was needed for defining the one word descriptors and several reviewers felt guidance should be provided as to how often and/or how rarely a "top score" should be given.

In the instruction manual for reviewers, a graphic was included to illustrate the relative balance of strengths and weakness associated with each rating score.   Responses indicated the majority (82%) of reviewers found the "strength/weaknesses graphic" at least somewhat helpful.   Reviewers commented that either no change was needed

for the strengths/weakness graphic or that the graphic was not even necessary because the score "verbal descriptors" provided sufficient guidance for deciding the final score.

How helpful was the strengths/weaknesses graphic (in the right column of reviewer training manual attached to e-mail) in your determination of the initial and final scores?
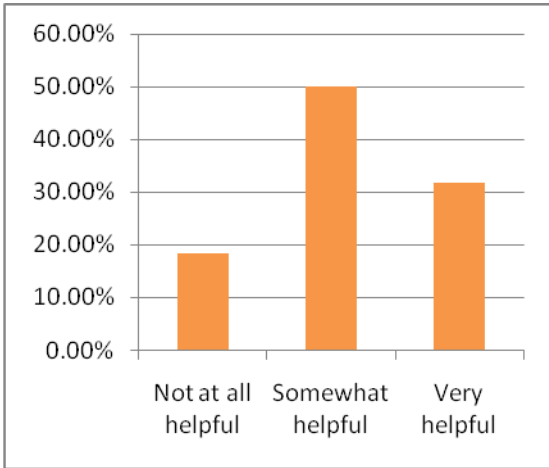

The evaluation survey contained four questions pertaining directly to the 1-9 scoring system itself. Responses indicated reviewers felt the new scoring system had a positive effect on their ability to determine the overall potential impact of applications reviewed. Reviewers also indicated the scoring system should have a positive effect on providing applicants an improved interpretation of written comments. Compared to the old 1-5 scoring system, reviewers in the pilot meetings with the 1-9 scoring felt the new scoring did not result in additional burden. The "overall level of satisfaction" with the new system was reported to be "very satisfied" by 68% of the reviewers and at least "somewhat satisfied" by the remaining 32%.

What effect do you feel the scoring of the specific criteria had on helping you determine the overall impact (merit) of the applications reviewed?
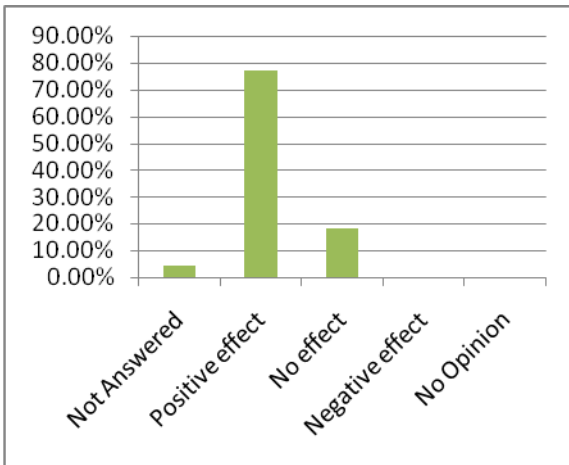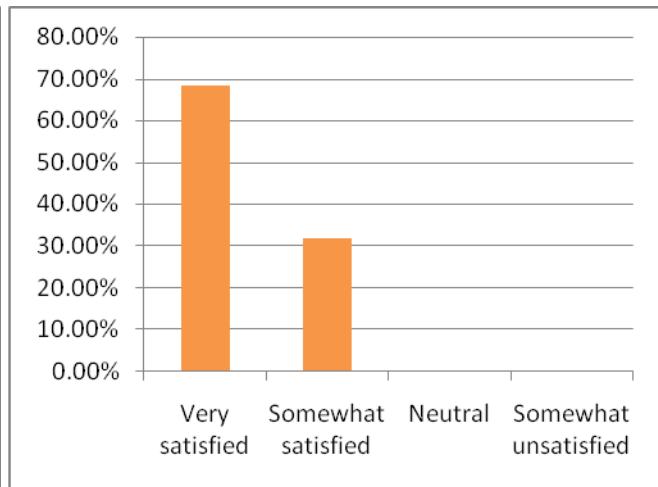
To what extent do you think having a numerical score for each of the five core criteria aids applicants' interpretation of written comments?
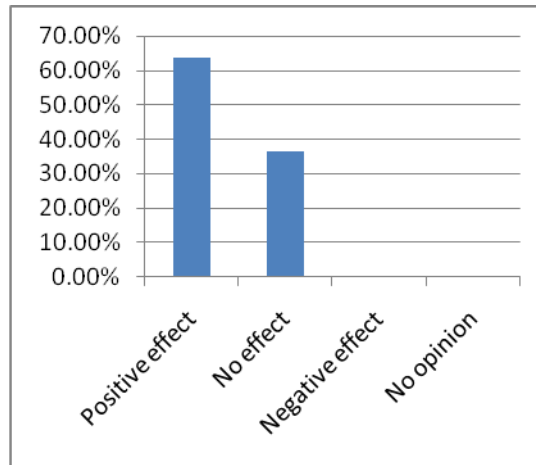
How much additional burden was the new system compared to the old scoring system?
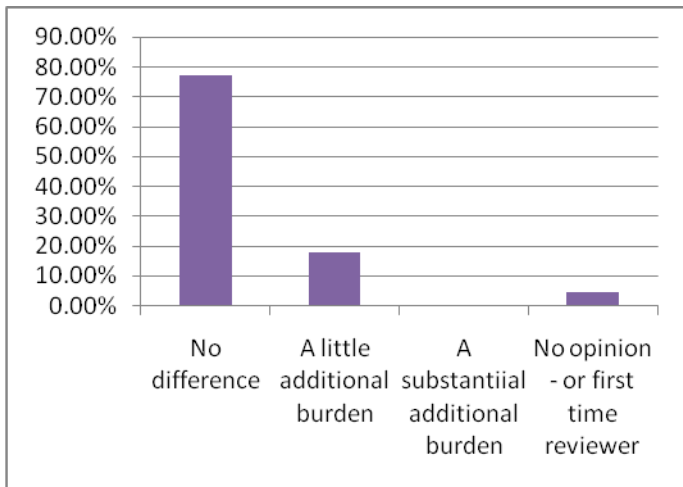
What is your overall level of satisfaction with a peer review process that includes the 1-9 scoring scale?

Reviewers were asked to rate the usefulness of the training materials they received. The majority (68%) reported that the materials were either "very useful" or "somewhat useful." Several open ended comments suggested a need to train reviewers in how to use the "full" 1-9 scale.

What is your overall level of satisfaction with a peer review process that includes the 1-9 scoring scale?
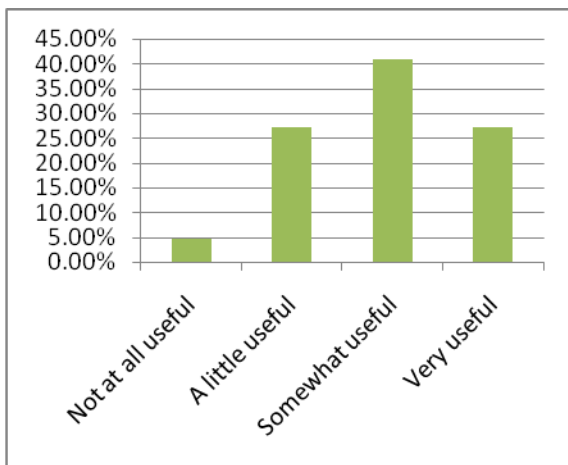
The final evaluation question asked, "What one thing would you change regarding the peer review process using the 1-9 scoring scale?" Two reviewers' comments pertained to the "weight" of each "criteria" in the final score and a need for more instruction to provide a better sense of how that process should occur. Other comments received pertained to the use of the scale. One reviewer stated, "we were advised to use a relative scale for scoring. However, the descriptors of the scores are in ABSOLUTE TERMS." In general, comments received revealed satisfaction with the new 1-9 scoring system as it is – with no modifications.