Strategies & Lessons Learned from Implementing External Peer Review Panels Online: A Case Example from a National Research Center

Presented by: Kelly Robertson
Team members: Daniela Schröter, Richard Zinser, Chris Coryn, & Pedro Mateu
The Evaluation Center, Western Michigan University
Purpose of Project

- Government Performance and Results Act (GPRA) measures:
  - Relevance of the research to practice
  - Quality of disseminated products

- 5 panel studies will be held
  - 2 Face-to-face
  - 3 Online
Panel Study

STRATEGIES & LESSONS LEARNED
Panelist Selection
Initial

1. E-mailed letter to invite nominations for panel
2. Nominees contacted and sign conflict of interest statement
3. Selection based on criteria:
   a) Expertise in field
   b) Outside field (e.g., evaluator or researcher)
   c) Prior experience on national panel
Panelist Selection
2nd Iteration and Beyond

1. Availability
2. Conflict of interest
3. Budget
4. Past performance
5. Participation in initial training
Panelist Selection

Lessons Learned

- Back-up panelists
- Diversity of panelists
Panel Logistics
Initial

- Face-to-face
- Individual ratings completed on-site
- All members rated all items ($n=2$)
- Panelist feedback survey
  - Logistics
  - Process
  - Strengths
  - Areas for improvement
# Panel Logistics

## 2\textsuperscript{nd} Iteration and Beyond

<table>
<thead>
<tr>
<th></th>
<th>2\textsuperscript{nd}</th>
<th>3\textsuperscript{rd}</th>
<th>4\textsuperscript{th}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items Reviewed</td>
<td>52</td>
<td>85</td>
<td>70</td>
</tr>
<tr>
<td># Panel subgroups</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td># Panelists per group</td>
<td>4</td>
<td>3-4</td>
<td>3</td>
</tr>
</tbody>
</table>
Panel Logistics
Lessons Learned

- Face-to-face meeting good for building rapport
- Included picture cheat sheet during virtual panel study
Training of Judges
Initial

- Introduced nature and intent of study
- Provided detailed instructions for using evaluation instruments
- Worked through hypothetical case
Training of Judges

2nd Iteration and Beyond/Lessons Learned

- Instructions provided but no training on rating
- Training on virtual software annually
Calibration of Judges
Initial

• Training: Hypothetical case
• Panel study:
  – Independent ratings
  – Discussion
  – Re-rating
Calibration of Judges

2nd Iteration and Beyond/Lessons Learned

- Clarification of rating scales by developing precise categories

<table>
<thead>
<tr>
<th>NA</th>
<th>Inadequate; fatal deficiencies</th>
<th>Barely adequate; some serious deficiencies</th>
<th>Adequate; minor, but nonfatal, deficiencies</th>
<th>Strong overall, but not exemplary; no deficiencies of consequence</th>
<th>Excellent overall; no deficiencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
</tbody>
</table>

- Discussion of individual ratings to overcome discrepancies vs group average
  - Diversity of raters
  - Error (e.g., panelists may miss information)
Independent Ratings

- **Initial:**
  - Conducted on-site at EC
- **2\textsuperscript{nd} Iteration and Beyond:**
  - Online
Independent Ratings
Lessons Learned

- Checklist
- Requiring justifications
- Timelines: 1-2 weeks between submission of ratings & panel
  - Multiple deadlines and reminders
Consensus Seeking

Initial

1. Group split into 2 subpanels and asked to rerate one study.
2. 2 subpanels came together to discuss new ratings and rationales.
3. Group deliberated on subpanel ratings and determined final ratings for each study.
Consensus Seeking
2\textsuperscript{nd} Iteration and Beyond

Group review all independent ratings during virtual panel meeting

Item Reviewed Title
Author 1 & Author 2

\(\bullet\) Panelist Name \(\bullet\) Panelist Name \(\bullet\) Panelist Name

<table>
<thead>
<tr>
<th></th>
<th>No merit; inadequate; fatal deficiencies</th>
<th>Limited levels of merit; weak; some serious deficiencies</th>
<th>Reasonably good; minor, but nonfatal; deficiencies</th>
<th>Very good, but not exemplary; no deficiencies of consequence</th>
<th>Excellent; exemplary; no deficiencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevance</td>
<td>NA</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>\textbf{Justification (Please indicate salient features, strengths, and weaknesses):} \textit{There obviously is a need for common data collection methods. So this has relevance for administrators and policymakers. The report is very relevant and addresses an issue that has the potential to cause harm to field. Strong need for commonality and longitudinal study.}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purpose</td>
<td>NA</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>\textbf{Justification (Please indicate salient features, strengths, and weaknesses):} \textit{It was an exploratory study, and the purpose was clear - to see what states are up to with respect to current data systems. The purpose was well stated. Yes, the report is clear about the danger of incomplete data and is clear about the proposed solution to this problem. I felt the purpose was very clear.}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(\Delta\) Average Per Item / Average for Proposal Overall = 8.1
Consensus Seeking  
Lessons Learned

Automated data visualization is time saving over alternative means

<table>
<thead>
<tr>
<th>Item Reviewed Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author 1 &amp; Author 2</td>
</tr>
</tbody>
</table>

| Panelist Name | Panelist Name | Panelist Name |

<table>
<thead>
<tr>
<th>Relevance: Relevance of product for intended audiences.</th>
<th>No merit; inadequate; fatal deficiencies</th>
<th>Limited levels of merit; weak; some serious deficiencies</th>
<th>Reasonably good; minor, but nonfatal, deficiencies</th>
<th>Very good, but not exemplary; no deficiencies of consequence</th>
<th>Excellent; exemplary; no deficiencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Justification (Please indicate salient features, strengths, and weaknesses):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>There obviously is a need for common data collection methods. So this has relevance for administrators and policymakers.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The report is very relevant and addresses an issue that has the potential to cause harm to field.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strong need for commonality and longitudinal study.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purpose: The degree to which the report has a clear purpose.</td>
<td>No purpose</td>
<td>Limited purpose</td>
<td>Reasonably clear purpose</td>
<td>Very good purpose</td>
<td>Excellent purpose</td>
</tr>
<tr>
<td>NA</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Justification (Please indicate salient features, strengths, and weaknesses):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It was an exploratory study, and the purpose was clear to see what states are up to with respect to current data systems. The purpose was well stated.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes, the report is clear about the danger of incomplete data and is clear about the proposed solution to this problem.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I felt the purpose was very clear.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

△ Average Per Item / Average for Proposal Overall = 8.1
Technology Used

- Planning:
  - WhenisGood

- Ratings:
  - Hosted Survey
  - Survey Monkey

- Virtual Panel Study
  - WebEx
Reporting of Findings

- Government wanted summative information but provided formative as well, as part of the panel.
  - Present results in aggregate across research proposals and product categories (summatively).
  - Presented results for each unique proposal and product (formatively) to allow researchers to improve their materials.
Audience

OTHER STRATEGIES?