Pilot Test Evaluation: Bias Mitigation in Peer Review Training for Program Staff

Health Research Alliance, July 2023

Summary

The Health Research Alliance (HRA) developed a Bias Mitigation in Peer Review Training for program staff at member organizations, which underwent a pilot test with a small group of participants in May and June 2023. The objectives of this pilot test were to collect formative qualitative feedback about the relevance and utility of the training content, the degree to which participants find the format engaging and accessible, recommendations for improving the training in future iterations, and to pilot survey questions that may be used to evaluate the effectiveness of subsequent versions.

Impact on Participants

The training had a positive impact on participants' confidence in their ability to identify bias and their comfort level with mitigating bias. Pre-training, most participants indicated that they did not often take action to mitigate bias during peer review. Post-training, most participants indicated that they intend to take action to mitigate bias during peer review.

Participants' Training Experiences

Overall, participants responded positively to the training. Participants found the training accessible, easy to navigate, comprehensive, and relevant for program staff at their organization. Participants found the case study, real-life examples, and specific and actionable strategies to mitigate bias during peer review to be particularly valuable. The focus on bias specifically in context of the peer review process and tailoring the content to the intended audience of program staff was appreciated by many participants. Many participants would welcome additional content in the form of more examples or supplemental resources (e.g., handouts, scripts for bias questioning, or other tools that can be referenced during a review meeting). Perceptions of the embedded reflection questions were mixed; some participants thought these would be more engaging or useful as multiple-choice questions instead of short-response, and some participants found the questions engaging or helpful as-is.

Suggested Revisions

1. Develop supplemental resources that participants can reference during or after the training.
2. Use survey/quiz questions more effectively to support engagement and allow participants to check their understanding.
3. Consider alternative ways to show bias mitigation strategies in action that do not require repeating long sections of the first case study.

Table of Contents

Training Description

The Health Research Alliance (HRA) Bias Mitigation in Peer Review Training is an asynchronous training for program staff at HRA member organizations. The training contained video presentations, reflection questions, and video case studies (total video runtime was approximately 30 minutes). The training was administered via Canvas Free for Teacher, a free version of the Canvas learning management system (LMS). The learning objectives for the training are:

1. Define bias, including types of bias and different manifestations of bias
2. Describe how bias can impact the review process
3. Identify strategies for mitigating bias during the review process

The content of the training is organized into three modules, one that addresses each of the learning objectives above.

Pilot Test Participants

Invitations to participate in the pilot test were sent via email to 32 individuals at 26 organizations who had previously expressed interest in learning more about a bias mitigation training module (largely via the November 2022 survey sent out by the Bias Reduction Training Working Group, or a smaller number who had verbally expressed interest at an HRA event). The invitation was also sent to members of the HRA Board of Directors (an additional 14 people who had not already expressed interest).

These invites yielded sign-ups from 40 individuals at 22 organizations. Of those who initially signed up to participate, 29 logged into the training and completed the pre-training survey, and 25 completed the training and post-training survey by June 15, 2023 (the stated end date of the pilot test window, after which the data were exported from the LMS and used for this evaluation). 15 people from 11 organizations also participated in a focus group after completing the training. According to pre-training survey responses, most of the participants were program staff (VP Programs, Program Director, Program Officer, etc.; $n = 16/29$, 55%) or operations staff (Grants/Contracts, Research/Evaluation, Communications, HR, IT; $n = 7/29$, 24%).

Table 1. Participants' role at their organization.

Role	# Respondents (% Total)
Program Staff (VP Programs, Program Director, Program Officer, etc.)	16 (55%)
Operations Staff (Grants/Contracts, Research/Evaluation, Communications, HR, IT, Other)	7 (24%)
Executive Officer (CEO, Exec. Dir, President, COO, CFO, etc.)	4 (14%)
Another role not listed	2 (7%)

Note: N = 29 responses.

Participants indicated which type of funding opportunities exist at their organizations. Nearly all participants said their organization offers research or established investigator grants (90%), and many offer research training and fellowships (52%) or career development awards (41%).

Table 2. Types of funding opportunities that exist at participants' organizations.

Funding Type	# Respondents (% Total)
Research or Established Investigator Grants	26 (90%)
Research Training and Fellowships	15 (52%)
Career Development Awards	12 (41%)
Program Project or Center Grants	4 (14%)
Resource Grants	2 (7%)
Another type of grant not listed	3 (10%)

Note: Respondents could select multiple options; percentages do not add up to 100%. *N* = 29 responses.

Impact on Participants

To determine the effect of the training on participants' ability to identify and mitigate bias, three questions were included on the pre- and post-training surveys. These questions query changes in participants' ability to identify bias, comfort level with taking action when bias is present, and their previous frequency and forecasted likelihood of taking action to mitigate bias. The questions were included to determine whether they will be useful in evaluating the effectiveness of future iterations of the training. All three of the questions on the pre-training survey yielded responses that were distributed across the range of response options, without concentration of many responses at the high end of the scales. This type of pre-training response skew would indicate that the questions are not structured in a way that pre- to post-training improvement can be effectively measured. These questions seem suitable for use in evaluation of training effectiveness, with the observation that more participants may select the 'I participate in peer review, but my role does not allow me to take action to mitigate bias' option on pre-training surveys than on post-training surveys (further described below, after *Figure 3*).

Participants' confidence in their ability to **identify** bias was higher post-training than pre-training (*Figure 1*). Pre-training, 21% of participants said they were very or extremely confident in their ability to identify bias during peer review; post-training, this increased to 62%.

How confident are you in your ability to identify bias during the peer review process?
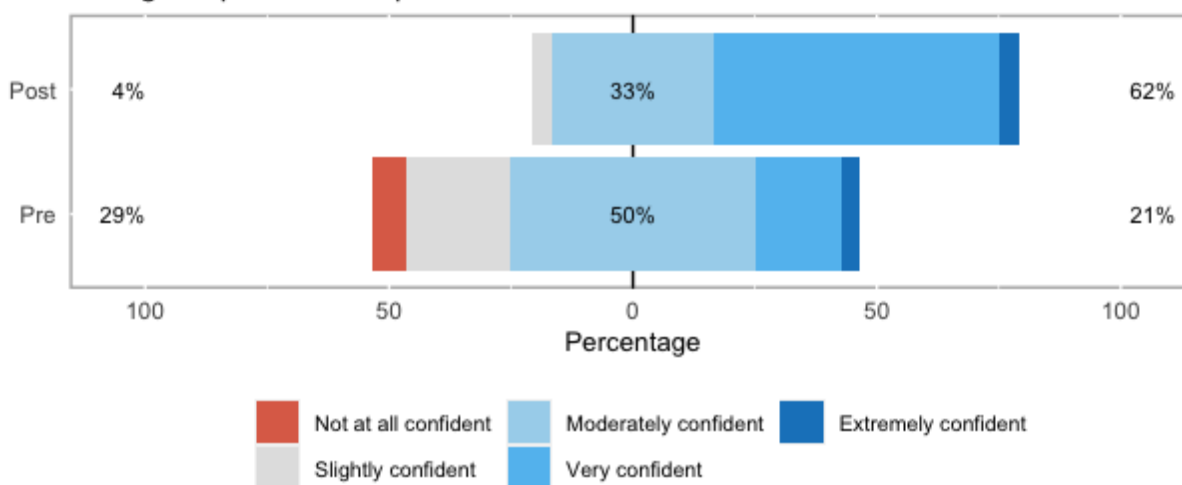

Figure 1. Participants' confidence in their ability to identify bias during peer review increased after taking the HRA bias mitigation training. Not shown in the graph: 1 pre-training response from a participant that indicated they do not participate in the peer review process. No respondents on the post-training survey chose this response option. *N* = 29 pre-training responses, *N* = 25 post-training responses. Percentages shown on the left, center, and right of the graph indicate the percentage of respondents who selected Not at all OR slightly confident, moderately confident, or very OR extremely confident, respectively.

Similarly, participants indicated a higher **comfort level with taking action** when bias may be present post-training compared to pre-training (*Figure 2*). Twice as many respondents rated their comfort level with taking action when bias may be present as 'very' or 'extremely' comfortable post- training (50%) compared to pre- training (25%).

How comfortable are you with taking action when bias may be present during the peer review process?
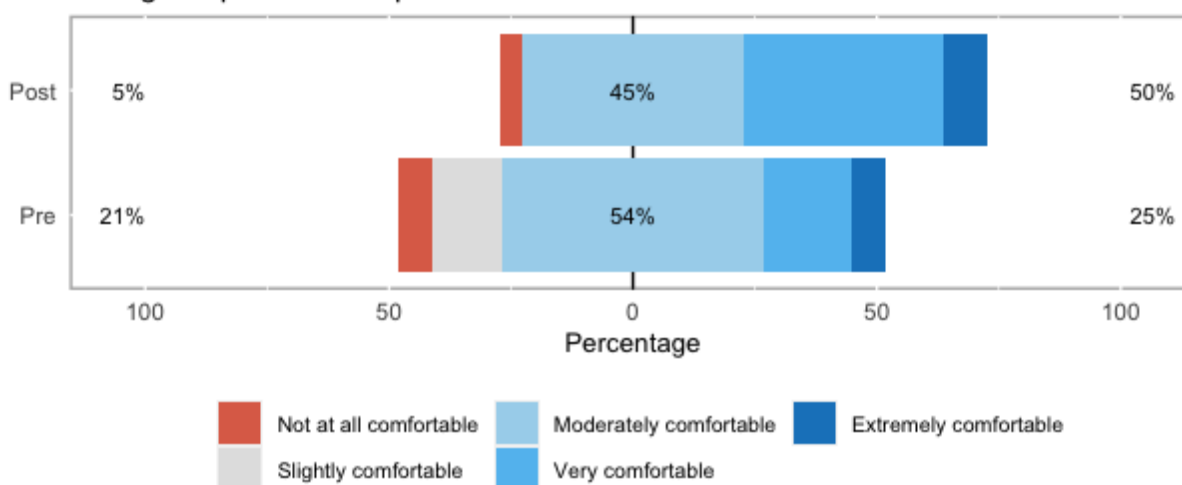
Figure 2. Participants' comfort level with taking action when bias may be present during peer review increased after taking the HRA bias mitigation training. Not shown in the graph: 1 pre-training response from a participant that indicated they do not participate in the peer review process, and 2 post-training responses from participants that indicated that they do participate in the peer review process, but their role does not allow them to take action to mitigate bias. Data shown on the graph include *n* = 28 pre-training responses, *n* = 23 post-training responses. Percentages shown on the left, center, and right of the graph indicate the percentage of respondents who selected Not at all OR slightly comfortable, moderately comfortable, or very OR extremely comfortable, respectively.

The pre-training survey asked participants how **frequently they took action in the past** to mitigate bias they noticed during peer review in the last year (*Figure 3*). Most respondents indicated that they never or seldom took action (55%, combined), and 20% indicated they usually or always took action. The post-training survey asked participants to forecast their future behavior by indicating **how likely they are to take action in the future** to mitigate bias in the future. Participants' responses suggest a strong positive intention to take action to mitigate bias; 88% said they will probably or definitely take action, and zero participants said they would definitely or probably *not* take action. While participants' responses on these two related, but different, questions cannot be directly compared, there seems to be a shift from a majority of participants tending towards *inaction* before the training to a majority of participants having a strong intention to *take action* to mitigate bias after the training.
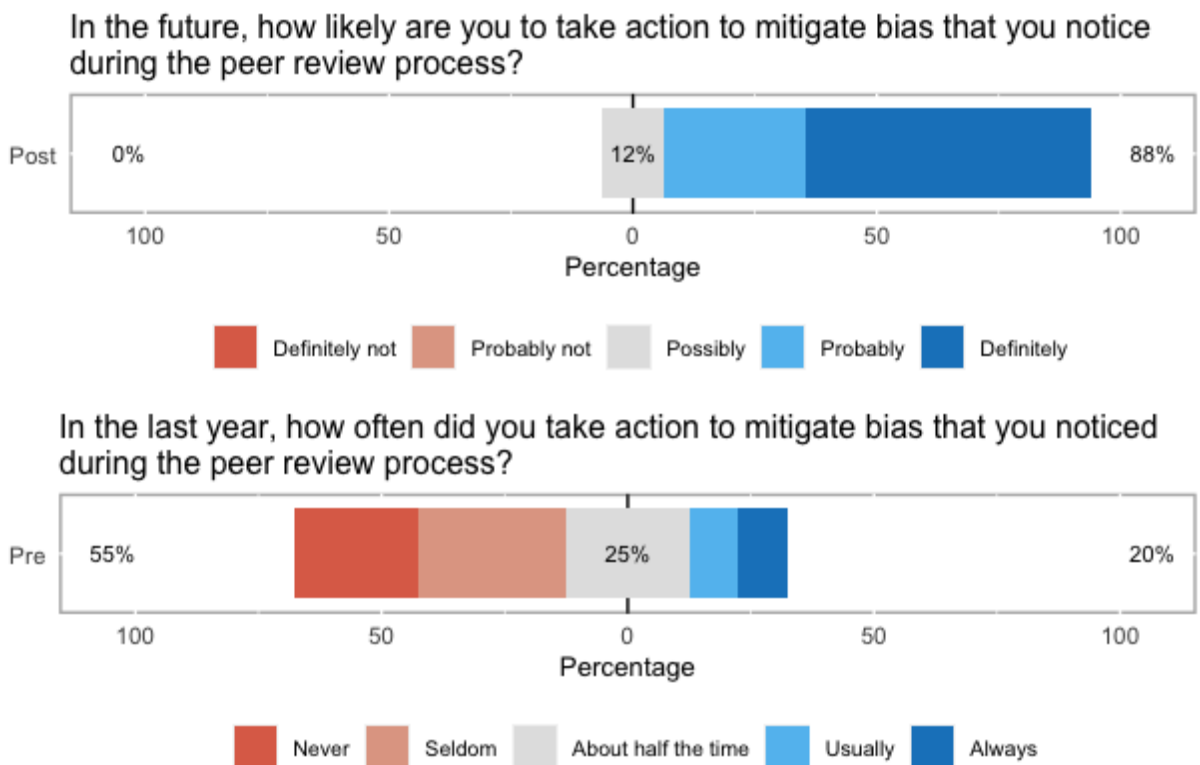

Figure 3. Participants may be more likely to take action to mitigate bias during peer review after taking the HRA bias mitigation training. Not shown in the graph: 9 pre-training responses from participants that indicated they do not participate in the peer review process (*n* = 4) or that they

participate in the peer review process but their role does not allow them to take action to mitigate bias (n = 5), and 1 post-training response from a participant that indicated that they do participate in the peer review process, but their role does not allow them to take action to mitigate bias. Data shown on the graph include n = 20 pre-training responses (bottom), n = 24 post-training responses (top). Percentages shown on the left, center, and right of the graph indicate the percentage of respondents who selected Never/Seldom (top) or Definitely not/Probably not (bottom), About half the time (top) or Possibly (bottom), or Usually/Always (top) or Probably/Definitely (bottom), respectively.

Both of these pre- and post-training survey questions included answer options for 'I did/do not participate in the peer review process' and 'I participate(d) in review sessions but my role did/does not allow me to take action to mitigate bias.' These responses are excluded from the data shown in *Figure 3*. On the pre-training survey, 4 participants indicated that they don't participate in peer review, and 5 participants indicated that they participate in peer review but were not allowed to take action to mitigate bias. On the post-training survey, only 1 participant indicated that they participate in review sessions but are not allowed to take action to mitigate bias (and no participants indicated that they don't participate in peer review). This change in responses could be due to participants not completing the training and thus only completing the pre-test survey. However, the sample size for the post-training survey is only 4 fewer than the sample size for the pre-training survey, so this cannot account entirely for the shift in responses. Another possible explanation is that the training provided examples of ways in which participants could take action to mitigate bias that they were not aware of previously, thus expanding participants' understanding of what 'taking action' means. The training prompted at least one participant to consider what is and what is not within their role description, as evidenced by this comment on the post-training survey:

> *"I think my comment is that I participate and think my role does not allow me to mitigate bias in the peer review discussion. I do however share comments from other reviewers with the group. I would let discussion play out among the peer reviewers rather than insert my thoughts. I feel more empowered to do so due to this training. I also see where our peer reviewers were doing this in our discussion. I guess I thought it was not my role to interfere. I do not really have clarity on that part of my role. I will discuss this further with our committee as I would also like to offer a brief training like this to our people."*

Participants' Training Experiences

Participants were asked on the post-training survey to indicate their agreement with the statements below (responses shown in *Figure 4*):

1. The training was accessible.
2. The training provided appropriate interventions or strategies to mitigate bias during a review meeting.
3. I am prepared to use what I learned in this training at my next peer review session.
4. The training was easy to navigate.
5. The training presented information that was relevant for program staff at my organization.

6. The information in the training addressed all of the topics I think are most important regarding bias in peer review.
7. The training effectively engaged me as a participant.

Overall, participants responded positively to the training and largely agreed with the statements above. Participants found the training accessible (100% agreed or strongly agreed), easy to navigate (96% agreed or strongly agreed), comprehensively addressed important topics (92% agreed or strongly agreed), and relevant for program staff at their organization (96% agreed or strongly agreed). Participants largely agreed that the strategies to mitigate bias presented during the training were appropriate (96% agreed or strongly agreed), and that they were prepared to use what they learned in the training at their next peer review session (96% agreed or strongly agreed). While many participants found the training engaging (76% agreed or strongly agreed), 8% disagreed that the training effectively engaged them as a participant and 16% neither agreed nor disagreed. This suggests that revisions to the training should aim towards improving how the training engages participants.
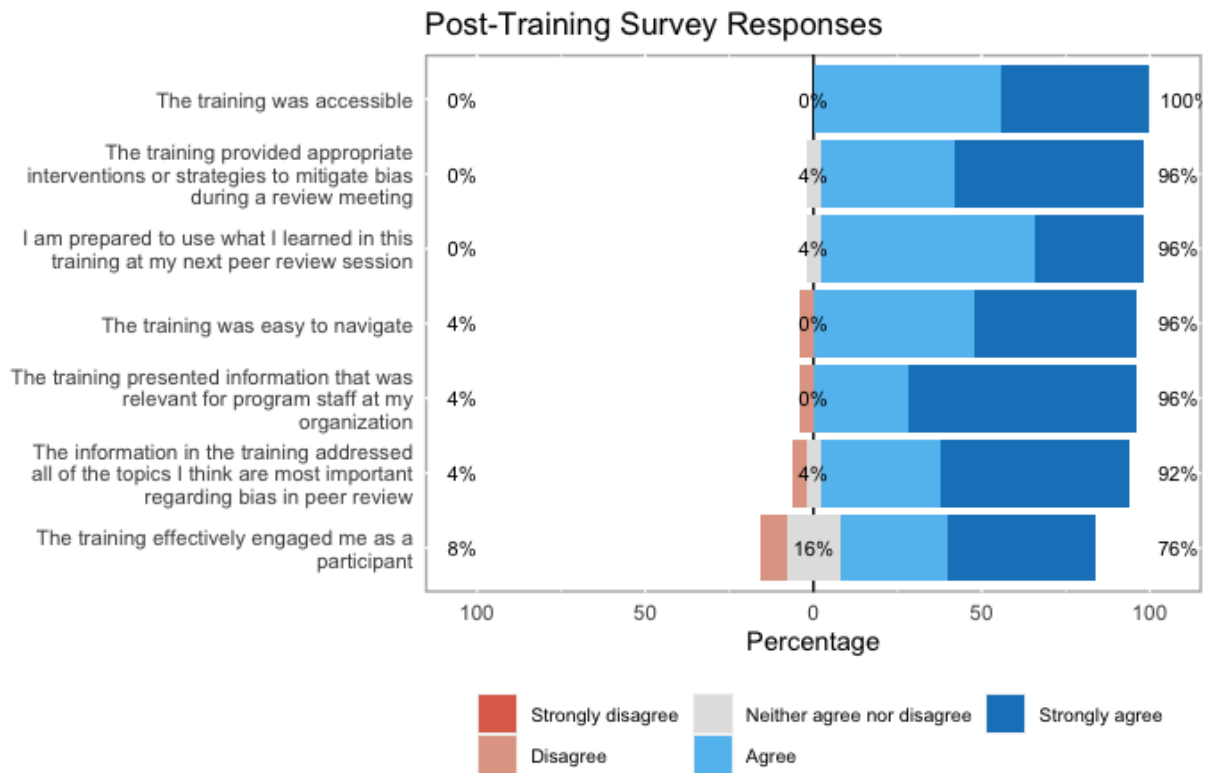

Figure 4. Participants post-training agreement with statements about the training. N = 25 post-training responses. Percentages shown on the left, center, and right of the graph indicate the percentage of respondents who strongly disagreed OR disagreed, neither agreed nor disagreed, and agreed OR strongly agreed, respectively.

The post-training survey included multiple short answer questions that asked participants about their experiences with the training; namely, what supported their learning and what was a barrier to their learning.

Supportive Factors

Content

Participants detailed factors that supported their learning in response to the post-training survey question "What element(s) of this training supported your learning? Why did you find this/these element(s) supportive?". Responses were inductively coded and categorized to surface common themes across responses. Participants discussed the content of the training (including the case study, examples, and strategies), as well as the format of the training.

The most common factor that participants cited as supporting their learning was the **case study** (or case studies). 15 of 25 participants referenced the case study(ies) in response to this question. Some responses referred to the case study in brief without elaborating on why this was supportive (e.g., "I liked the case scenarios," or "I really appreciated the case studies and questions.") or used positive but vague adjectives to describe the case studies ("interesting and fun" or "useful and great"). Some responses affirmed that the case study scenarios were true to their real-life experiences, and that seeing examples of implementing the strategies was helpful. The focus groups also echoed the sentiment that the case studies contained realistic examples and were effective in demonstrating how bias manifests during peer review and how to implement the strategies discussed in the training. Example comments from the post-training survey about the case studies:

> *"I really liked watching the two case studies and how the individuals put into action potential bias mitigators. It made it seem less 'intimidating' to use this on a review committee."*

> *"I liked how there was an actual example of how a review call goes and the comments that were made that accurately represented that different types of bias that were mentioned."*

> *"The case study was the most helpful. I've done a variety of different antibias trainings and I always find those that dive into specific examples the most helpful. I think it's always helpful to see concepts in a more applied manner."*

In addition to the case study(ies), other content components in the training were identified as supportive factors by 11 participants. Most of these responses referred to either the **examples** included in the training (6 responses), or the **strategies** presented to mitigate bias during peer review (5 responses). The "examples" referenced in the responses could include the examples of different types of bias presented in Module 2, examples of statements within the case studies, or example questioning statements to use to mitigate bias during peer review (presented in Module 3). It was often not possible to discern which type of examples to which participants referred based on their responses. Positive attributes of the examples that participants described included that they were relevant, realistic, and informative. Participants described the strategies as easy to follow or implement, helpful, realistic, actionable and specific. This sentiment

was also validated by the responses from the focus group participants, many of whom found the strategies practical, reasonable to implement, and well-tailored to the specified audience of program staff. Examples of comments about the examples or strategies presented in the training:

> *"This training provided reminders of many of the non-application specific topics that get discussed in study section. But more importantly, introducing actionable steps we can take to mitigate bias, rather than solely recognizing was very helpful and will make this easy to implement."*

> *"The examples were very relevant and the suggestions to combat bias were easy to follow."*

Additional content elements participants identified as supportive included the summary page at the end of the training (1 response), definitions (1 response), the overview of common biases (1 response) and the graphic organizer (1 response; likely referring to the Cognitive Bias Codex linked in Module 2).

Format
The format of the training was referenced as a supportive factor by 5 participants. 3 of these participants referred to the embedded **reflection/assessment questions** that followed the case studies as being supportive. One participant said the questions helped them remain engaged, and the other participants used positive but vague descriptors when referencing the questions (e.g., "extremely helpful" and "I appreciated the case studies and questions"). In one of the focus groups, multiple participants stated that they found the open-text reflection questions engaging and appreciated that they prompted participants to deeply reflect on the material and to cohesively articulate their thoughts.

Other format elements that were identified as supportive in both survey responses and in the focus groups were the **length** of the training and the ability to read **captions** on the videos.

Unsupportive Factors
There were two questions on the post-training survey that queried participants' negative experiences with the training: "What barriers, if any, made participation or engagement difficult?" and "What element(s) of this training did NOT support your learning? Why did you find this/these element(s) unsupportive?". In this analysis, the responses for these questions were combined, as the two questions yielded similar responses for many participants. This suggests that it is not necessary to ask both questions on subsequent surveys. Responses were inductively coded and categorized to surface common themes across responses.

9 participants did not note any barriers or unsupportive factors in either of the questions. Some responses included positive comments (e.g., "no aspects were unsupportive. I found it short and educational which is a great combination"), and some responses

provided no additional information (e.g., "N/A", "None", or a blank response with no text).

Content
8 participants identified some aspect of the content presented in the training as something that did not support their learning and/or acted as a barrier to their participation. Three participants suggested including **additional resources** like summary sheets, a list of types of biases, sample scoring guides, or handouts designed for use during a review meeting to make it easier for participants to remember and implement key takeaways from the training. A desire for additional resources was also discussed by all of the focus groups, with participants noting that quick reference sheets, listening primers, scripts, or other tools would support learning during the training experience and implementation of strategies after the training has ended.

Two participants noted that existing content could be explored in **more depth**, either by citing additional references or by including additional opportunities to practice the concepts in action. Two participants referenced that the **second case study was repetitive** in content, and the format of repeating the parts from the first case study that did not change made it difficult to pull out the new information, particularly because of the increased length of the second case study. This sentiment was also discussed in one of the focus groups, in which participants noted that the second case study video was long, at the end of the training, and repetitive, making it difficult to keep focus and find the useful information within. One survey response said that the examples of biases presented in the case study were too obvious, and that they think "most review committees have moved past the obvious kinds of bias by now". One participant noted that some organizations may not include review meetings as part of their peer review process and suggested to broaden the content such that it's generalizable across a wider range of funding processes. Example quotes that relate to the content of the training as being a barrier or otherwise not supportive:

> [Referring to the second case study video] *"Again, repeating the same video with a few examples of how to address bias. This wasn't very helpful because the nuggets of useful information are hidden inside a video that the participant has already listened to. Maybe there is a better way of doing this. Maybe just showing short clips after specific comments are made instead of replaying the video completely."*

> *"I would have retained more if summary "cheat sheets" were provided alongside the video as references to help me retain information and provided at the end to take it back to my normal life."*

Format
Most of the participants who specified a barrier or unsupportive factor referred to something about the format or logistics of the training (14 participants). These factors included:

- The **reflection questions** (7 participants). Comments largely focused on the type of question structure used (5 participants). Both questions were open-text questions; some participants suggested using multiple choice questions to improve engagement and/or interactivity, to make questions easier to respond to, or to allow participants to confirm that they understand the concepts in the training. Some participants noted some LMS artifacts that were confusing or otherwise negatively impacted their experience, like the message saying 'correct responses are hidden' after submission (which is not relevant for a survey with no correct responses), or that the question and response submission box were not available while the participant was watching the video, making it difficult to remember information that they then wanted to write in response to the question after several minutes of video. One of the focus groups discussed ways in which the reflection questions could be improved, including explicit instructions that the activity contains a short-answer questions, listening guides to prime participants to notice key points or take notes, or other referenceable materials (e.g., the additional resources referenced above).
- The **video design and/or video logistics** (6 participants). Two participants said that the *slide design* was unsupportive or a barrier, citing either the simple graphics or the balance of text to graphics as elements that made engagement difficult to sustain. Two participants mentioned a need for *captions* (which were present on each video and noted in the text above each video). Two participants referenced that the *voiceover* was not optimal (one participant found the voiceover robotic which negatively impacted their focus, and also found the speaking speed too slow; the other participant thought the voiceover may be too fast if the content is new information for the viewer/listener).
- The **timing** of the training (2 participants). One participant indicated they would like to see more time spent on the different types of bias, and one participant said the training should be shortened so it is no more than 30 minutes in length.
- How to **navigate** the LMS (1 participant). One participant noted they had difficulty determining how to navigate to the next video.

Relevance to Participants

Nearly all participants (96%) agreed or strongly agreed that the training is relevant for program staff at their organizations (*Figure 4*). Multiple focus group participants referenced the clear relevance to program staff at funding organizations who deal with peer review as being a key point of unique value that this HRA training provides. Participants also indicated their response to the question "This training would be relevant for program staff at my organization who work on the following types of funding mechanisms," and the response data is shown in *Table 3*. The most frequently selected funding mechanisms were Research or Established Investigator Grants (88% of respondents) and Research Training and Fellowships (72% of respondents). This aligns with the pre-survey data in which participants indicated which types of funding exist at their organization; these two mechanisms were the most common. The percentage of participants who thought the training would be relevant for staff who work on Resource Grants or Program Project or Center grants were higher than the percentage of participants who indicated that these opportunities exist at their organization.

Table 3. Participants' perception of training relevance for program staff working on various funding mechanisms.

Funding Type	# Respondents (% Total)
Research or Established Investigator Grants	22 (88%)
Research Training and Fellowships	18 (72%)
Resource Grants	9 (36%)
Program Project or Center Grants	7 (28%)
Another type of grant not listed	5 (20%)

Note: Respondents could select multiple options; thus, percentages do not add up to 100%. *N =* 25 responses. The response option 'Career Development Awards' was included on a similar question about funding mechanisms in the pre-training survey, but mistakenly omitted from the post-training survey.

Suitability of the Training for Reviewers

This iteration of the training was developed specifically for program staff at HRA member organizations. Participants were asked whether a similar training would be useful for peer reviewers at HRA member organizations, and all of the participants indicated that a similar training for peer reviewers would be useful, at least to some degree (*Figure 5*). 88% of participants thought a similar training would be extremely or very useful.
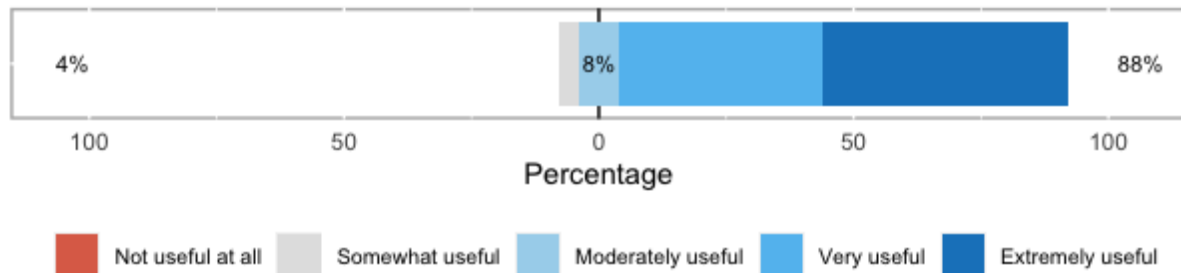

Figure 5. Participants believe a similar training for peer reviewers would be useful. *N* = 25 post-training responses. Percentages shown on the left, center, and right of the graph indicate the percentage of respondents who selected 'not useful at all' OR 'somewhat useful, 'moderately useful', or 'very useful' OR 'extremely useful', respectively.

Participants were asked "How would you recommend modifying this training if HRA were to design a training for peer reviewers?" Their responses were inductively coded and analyzed for common themes, which are discussed below.

Content
16 participants recommended modifying an element of the training content to tailor it to an audience of peer reviewers. 11 participants suggested additional content elements that could be included in a training designed for peer reviewers:

- **Additional resources** (4 participants), including a handout describing the mitigation strategies, sample scoring guide, a "script" for statements that can be used to intervene when bias is noticed, and tools or resources that can be referenced during a meeting to help reviewers employ strategies covered in the training. Participants also indicated that supplemental resources would be supportive for program staff.
- **Additional activities or topics** (5 participants), including writing non-biased critiques, mock study section activity, a case statement open discussion, using bias mitigation techniques during individual review, sharing examples of bias and what they did to mitigate them, and opportunities to practice strategies or apply concepts. One participant recommended including additional types of content, like op-eds.
- **Explore existing content with additional depth** (3 participants), including presenting data and including more detail about the research on bias and include additional examples or case studies showing different types of bias.

6 participants suggested revising or reframing existing content, such as:
- **The case study** (3 participants), including having short examples that each cover one type of bias instead of one scenario that includes many types; ensuring that the case study is sufficiently distinct from the case study presented in the NIH bias training (with which many reviewers are already familiar); and, more generally, breaking the case study into smaller parts.
- **The strategies** to mitigate bias (2 participants), including removing this section or reframing this section to focus on peer-to-peer bias mitigation strategies.
- **Broadening the focus** (1 participant) to encompass a wider range of review processes that may not all include face-to-face peer review meetings.

Format

6 participants recommended modifying one or more format elements to tailor the training to an audience of peer reviewers. These format elements include:
- Using **fixed-response questions** instead of short answer questions for self-assessment (2 participants)
- Decreasing the length (2 participants)
- Ensuring **captions** are available (1 participant; captions were available for all videos)
- Combining all the content into a **single video** (1 participant)

Some of the comments discussed in the 'Content' section above also indicate a desire for more varied types of formats, including activities where participants can practice strategies in action (5 participants).

No changes needed

5 participants indicated that no changes were necessary to make this training useful for peer reviewers; these participants thought the training would be helpful, useful, or effective as-is.

Suggested Revisions (for Program Staff Training)

Participants' feedback on the training suggest the following revisions to the training for program staff, which were themes summarized across multiple participants' responses:

1. Develop supplemental resources that participants can reference during or after the training, such as:
   a. Definitions of different types of bias
   b. Example statements that indicate different types of bias
   c. Scripts for bias questioning statements
   d. Lists of strategies to implement during review meetings
2. Use survey/quiz questions more effectively to support engagement and allow participants to check their understanding. Potential revisions:
   a. Clarify the instructions about what is expected (e.g., participants may not have known that the survey contained a short-answer question because the instructions were vague)
   b. Include supplemental resources that make it easier to answer the question (e.g., a vocabulary bank that describes the different types of bias for participants to reference when reflecting on the case study)
   c. Use multiple-choice questions so participants can confirm whether their understanding of concepts is correct
   d. Minimize as many not-educationally-relevant Canvas LMS artifacts as possible, and where not possible, provide an explanatory note to minimize confusion (e.g., the "correct answers are hidden" message in response to survey questions that don't actually have a correct answer)
3. Consider alternative ways to demonstrate bias mitigation strategies in action that do not require repeating long sections of the first case study.
   a. Relatedly, consider technical improvements to the 'Check your knowledge' Google Slides activity, which some participants reported to be glitchy.

The following revisions could also be considered, which were informed by a smaller number of participants' comments than the three revisions listed above:

4. Modify the slides and/or graphics to contain more visuals and less text, and/or more sophisticated graphics.
   a. This could also include using video of real people instead of the cartoon images in the case study.
5. Generalize the content to apply to program staff that work in review processes that do not include a peer review meeting.
6. Create discussion resources or opportunities to facilitate peer learning around the bias mitigation training. This could include discussion guides to be used by HRA members at their own organizations or discussion sessions to facilitate cross-organization information exchange.

Any revisions made should not increase the length of the training, as participants largely valued the brief nature of the training (or advocated for it to be shorter).