

Minnesota State Arts Board

FY2021 Rating Criteria Reliability and the Effects of Equitable Funding Strategies

Introduction

From 2020-2021, Wilder Research (Wilder) assisted the Minnesota State Arts Board (MSAB) in assessing the reliability of its review criteria and the effects of its equity funding strategies. MSAB provided reviewer ratings for the FY2021 Creative Support grant programs to Wilder, along with an explanation of its equitable funding strategies for three priority groups: applicants who are Black, Indigenous, or a person of color (BIPOC); applicants with disabilities; and applicants from greater Minnesota.

Recommendations

Based on the results of our criteria reliability testing and analysis of the effects of MSAB's equitable funding strategies, we recommend the following actions for MSAB to consider:

- **Explicitly communicate to priority group applicants to include their priority group identification in their application narrative.** For the equity-focused rating criteria, BIPOC or BIPOC-led applicants on average received higher scores than non-priority applicants (the intended purpose of the equity-focused criteria). However, this was not true for the greater Minnesota applicants or for the disability priority group. We expect this may be because these applicants did not identify themselves as a priority group (e.g., disabled, from greater Minnesota) in their application narrative and, as a result, reviewers did not know to rate their application as such. See pages 5 & 9 for the findings that catalyzed this recommendation.
- **Explicitly communicate all priority groups to reviewers—and that identification or connection with just one group merits a high-level rating for equity-focused criteria.** Similar to the previous recommendation, this recommendation was catalyzed by the findings on pages 5 & 9—namely, that the equity-focused criteria worked “as intended” for BIPOC and BIPOC-led applicants, but not for the other priority groups. This may be because reviewers did not take all priority groups into consideration when rating applicants or because reviewers subjectively prioritized BIPOC applicants over other priority group applicants.

- **Revise certain applicant rating criteria to increase reliability (based on reliability testing results).** Our reliability testing showed low reliability for some rating criteria (and the equity-focused criteria in particular). We recommend that MSAB revise these criteria to increase reliability for future grant cycles. See pages 4 & 7-8 for the findings that catalyzed this recommendation, as well as the Appendix for specific sections of the criteria that might improve reliability if updated.

Methods

Testing the reliability of review criteria

We assessed the reliability of MSAB’s review criteria overall and with a focus on the reliability of criteria for priority group applications. This analysis resulted in two measures of reliability:

- **An inter-reviewer reliability score.** This is a typical approach for measuring reliability. It measures how consistently multiple reviewers employed review criteria for the same application.
- **A reviewer agreement score.** We used this approach when the majority of reviewers used the same rating (e.g., the highest rating, “6”) for a particular criterion. The statistical process for reliability testing assumes a certain degree of variability in ratings (e.g., a “good spread” of ratings); if there is no variability, this is an issue for the statistical test. As such, we used this alternative approach when there was minimal variability in ratings.

Examining the effects of equity funding strategies

We examined the average rating received by priority group applicants as compared to non-priority group applicants for equity-focused criteria (criteria 3a-3c). This allowed us to determine the extent to which these criteria were working as intended—that is, intentionally elevating the ratings for these priority group applicants because historically these groups have been underserved by MSAB.

In addition to examining the effect of equity-focused rating criteria for priority group applicants, we also assessed the degree to which MSAB’s funding allocation approach helped or hindered its equity goals (that is, providing a higher proportion of funding as compared to previous years to priority groups). The funding allocation approach used by MSAB included the following steps:

1. Determine the total percentage of all applicants that can be funded considering the amount of funds available for each grant program (for the Creative Support for Individuals program, for instance, 79% of all applicants could be funded)

2. Award that percentage of applicants from each priority group (for the Creative Support for Individuals program, for instance, the top 79% of BIPOC applicants were funded, the top 79% of disabled applicants were funded, and the top 79% of greater Minnesota applicants were funded)
3. Award as many remaining applicants as possible, considering the amount of funds still available (this is regardless of priority group status, so if a priority group applicant was “next in line” for general funding based on their ratings, they would be awarded a grant)

To determine the effect of this funding allocation approach in regards to MSAB’s equity goals, we modeled two comparative allocation approaches: 1) allocating funds based on total scores only, regardless of whether applicants identified with priority groups and 2) allocating funds to match the proportion of priority group applicants in the overall applicant pool. For both Creative Support programs (Individuals and Organizations), we compare the makeup of successful applicants based on these three funding allocating approaches.

Creative Support for Individuals

Creative Support for Individuals was designed to help Minnesota artists and culture bearers adapt to changes in their working environment caused by the global pandemic. Grantees were able to use funds to sustain their practice and stay relevant and connected to audiences, participants, students, or communities now and in the future. This grant program aimed to help Minnesota artists and culture bearers maintain their visibility and financial sustainability by using their creativity and connections to community.

Application rating criteria reliability

Please see Figure 1 for a summary of the results from our reliability testing of this program’s review criteria in FY2021. We assigned three different levels of reliability for each criterion—higher, medium, and lower. High reliability refers to a reliability score of 0.60 or greater or an agreement score of 60% or greater. Medium reliability refers to a reliability score that is 0.30-0.59 or an agreement score that is 40-60%. Low reliability refers to a reliability score that is 0.29 or lower or an agreement score that is 29% or lower.¹ For each criterion, we note its reliability level and offer an explanation for why the criterion received this reliability level via our testing.

¹ Please note that these reliability levels differ somewhat from traditional reliability measures. For instance, a “strong” reliability level, traditionally, is 0.8-0.9; “moderate” is 0.6-0.79; “weak” is 0.4-0.59. We chose these breakpoints for “higher,” “medium,” and “lower” reliability levels because our testing results did not align well with traditional measures. This may be because the rating criteria used by MSAB (and the process by which applications are rated) differs in substantive ways from the criteria and processes for which reliability testing is typically employed as a method by which to assess reliability.

Please see the Appendix for specific sections of the criteria that may have contributed to low reliability levels.

1. Rating criteria reliability levels and explanations

Criteria	Reliability level	Explanation
1a	Higher	This criterion had an agreement score of 70% (high agreement among reviewers) because most reviewers used the highest rating (“6”). This criterion will likely be more useful (from a rating perspective) if it was updated so that fewer applicants received the highest rating. Additionally, rating “6” would benefit from defining “highly engaged,” “high number,” and “consistently shares and creates.”
1b	Higher	This criterion had an agreement score of 67% (high agreement among reviewers) because most reviewers used the highest rating (“6”). This criterion will likely be more useful (from a rating perspective) if it was updated so that fewer applicants received the highest rating.
2a	Lower	This criterion had a reliability score of 0.20 (low reliability across reviewers). This is likely because the explanations for possible ratings are not clear or distinct. In particular, we suggest that ratings “3” and “1” are not particularly distinct from each other.
2b	Lower	This criterion had a reliability score of 0.24 (low reliability across reviewers). This is likely because the explanations for possible ratings are not clear or distinct. In particular, we suggest that ratings “3” and “1” are not particularly distinct from each other.
3a	Higher	This criterion had an agreement score of 64% (high agreement among reviewers)—because the explanation for rating “4” is particularly clear (whether the applicant identifies with at least one priority group). That being said, we suggest that “deep connection” in rating “2” should be better defined.
3b	Medium	This criterion had an agreement score of 58% (medium agreement among reviewers). Similar to others, this criterion will likely be more useful (from a rating perspective) if it was updated so that fewer applicants received the highest rating. Additionally, the distinction between rating “4” and rating “2” is not particularly clear.
3c	Lower	This criterion had a reliability score of 0.25 (low reliability across reviewers). This is likely because the explanations for possible ratings are not clear or distinct. In particular, we suggest “frequent” and “sustained connections” in rating “4” should be better defined and that the distinction between rating “2” and rating “0” is made more clear.

Effects of equity-focused rating and funding allocation

The equity-focused rating criteria were intended to elevate the ratings of priority group applicants because historically these groups have been underserved by MSAB. These criteria include:

- 3a. Applicant identifies with underserved community/ies
- 3b. Significance of applicant’s work to underserved community/ies
- 3c. Applicant has a sustained connection to underserved community/ies

Figure 2 shows that these criteria worked as intended for BIPOC applicants (that is, BIPOC applicants received higher average ratings for these criteria as compared to non-priority applicants). For applicants with a disability, these criteria had no meaningful effect on average rating (that is, there was no meaningful difference in average ratings for these criteria between disabled applicants and non-priority group applicants). Lastly, for applicants from greater Minnesota, these criteria had the opposite effect from what was intended—these applicants received lower average ratings for these criteria as compared to non-priority applicants.

2. Equity-focused ratings for priority group applicants vs. non-priority group applicants

Criteria	BIPOC applicants	Applicants with a disability	Greater Minnesota applicants
3a. Applicant identifies with underserved community/ies	Higher than non-priority	No meaningful difference	Lower than non-priority
3b. Significance of the applicant's work to underserved community/ies	Higher than non-priority	No meaningful difference	Lower than non-priority
3c. Applicant has a sustained connection to underserved communities	Higher than non-priority	No meaningful difference	Lower than non-priority

Note. "Higher than non-priority" refers to when a priority group's average rating for these criteria was 10 or more percentage points *higher* than the non-priority group's average rating. Similarly, "lower than non-priority" refers to when a priority group's average rating for these criteria was 10 or more percentage points *lower* than the priority group's average rating.

Overall, 566 individuals applied for a Creative Support for Individuals grant. Of those, 237 applicants identified as BIPOC, 98 applicants identified as having a disability, 273 applicants resided in greater Minnesota (outside of the Twin Cities metro), and 83 were non-priority group applicants. Please note that these figures add to more than 566 because some applicants identified with more than one priority group. See Figure 3 for a breakdown of applicant type.

3. Percentage breakdown of applicant type

Applicant type	% of total applicants
BIPOC applicants	42%
Applicants with a disability	17%
Greater Minnesota applicants	48%
Non-priority group applicants	15%

Note. These percentages add to more than 100 percent because some applicants identified with multiple priority groups.

MSAB determined that 447 applicants (79%) could be funded with funds available through the Creative Support for Individuals grant program. First, MSAB awarded the top 79% of applicants from each priority group—187 BIPOC applicants, 76 disabled applicants, and 216 greater Minnesota applicants. After awarding these applicants, there were enough funds remaining to award the top rated 56 remaining applicants, regardless of whether they identified with a priority

group. Please see Figure 4 for the effects of this funding allocation and compares this to allocating funding by total score.

4. Percentage of successful applicants by applicant type for two funding allocation approaches

Applicant type	79% funding allocation	Total score funding allocation
BIPOC applicants	84% of applicants	85% of applicants
Applicants with a disability	79% of applicants	74% of applicants
Greater Minnesota applicants	79% of applicants	72% of applicants
Non-priority group applicants	68% of applicants	83% of applicants

Creative Support for Organizations

Creative Support for Organizations was designed to help arts organizations and non-arts organizations that regularly offer arts programming as an integral part of their mission adapt to the changing environment caused by the global pandemic. Grantees were able to determine how best to use funds to stay relevant and connected to the audiences, participants, students, or communities that participate in their regular arts programming. By using their creativity and connection to community, this grant program was intended to help Minnesota organizations maintain the long-term viability of their arts programming so that Minnesota residents and communities will maintain access and connection to the arts.

Application rating criteria reliability

Please see Figure 5 for a summary of the results from our reliability testing of this program’s review criteria in FY2021. We assigned three different levels of reliability for each criterion—higher, medium, and lower. High reliability refers to a reliability score of 0.60 or greater or an agreement score of 60% or greater. Medium reliability refers to a reliability score that is 0.30-0.59 or an agreement score that is 40-60%. Low reliability refers to a reliability score that is 0.29 or lower or an agreement score that is 29% or lower.² For each criterion, we note its reliability level and offer an explanation for why the criterion received this reliability level via our testing.

² Please note that these reliability levels differ somewhat from traditional reliability measures. For instance, a “strong” reliability level, traditionally, is 0.8-0.9; “moderate” reliability is 0.6-0.9; “weak” is 0.4-0.59. We chose the described breakpoints for “higher,” “medium,” and “lower” reliability levels because our testing results did not align well with traditional measures. This may be because the rating criteria used by MSAB (and the process by which it is implemented) differs in substantive ways from the typical criteria and processes for which reliability testing is employed as a method by which to assess reliability.

Please see the Appendix for specific sections of the criteria that may have contributed to low reliability levels.

5. Rating criteria reliability levels and explanations

Criteria	Reliability level	Explanation
1a	Higher	This criterion had an agreement score of 73% (high agreement among reviewers) because most reviewers used the highest rating (“6”). This criterion will likely be more useful (from a rating perspective) if it was updated so that fewer applicants received the highest rating. Additionally, rating “6” would benefit from defining “highly engaged,” “high number,” and “consistently shares and creates.”
1b	Higher	This criterion had an agreement score of 68% (high agreement among reviewers) because most reviewers used the highest rating (“6”). This criterion will likely be more useful (from a rating perspective) if it was updated so that fewer applicants received the highest rating.
2a	Higher	This criterion had an agreement score of 62% (high agreement among reviewers) because most reviewers used the highest rating (“6”). This criterion will likely be more useful (from a rating perspective) if it was updated so that fewer applicants received the highest rating. In particular, we suggest that ratings “3” and “1” are not particularly distinct from each other.
2b	Lower	This criterion had a reliability score of 0.18 (low reliability across reviewers). This is likely because the explanations for possible ratings are not clear or distinct. In particular, we suggest that ratings “3” and “1” are not particularly distinct from each other.
3a	Lower	This criterion had a reliability score of 0.24 (low reliability across reviewers). This is likely because the explanations for possible ratings are not clear or distinct. In particular, we suggest that “inclusive of and deeply connected to” in rating “4” could benefit from revisions to increase clarity. Similarly, we suggest that the either/or in rating “2” may present difficulties to reviewers—specifically in that the first option notes “no evidence that leadership roles are filled by members of underserved communities” but the second option notes “members of underserved communities serve on a permanent paid advisory council/committee.” One reading of these options is that serving on an advisory council is an example of leadership—and so these two options, in this reading, are contradictory.
3b	Lower	This criterion had a reliability score of 0.20 (low reliability across reviewers). This is likely because the explanations for possible ratings are not clear or distinct. In particular, we suggest that ratings “4” and “2” are not particularly distinct from each other.
3c	Lower	This criterion had a reliability score of 0.14 (low reliability across reviewers). This is likely because the explanations for possible ratings are not clear or distinct. In particular, we suggest that ratings “4” and “2” are not particularly distinct from each other. Further, we suggest that “frequent,” “intentional,” and “sustained connections” in rating “4” could benefit from clear definitions.

Effects of equity-focused rating and funding allocation

The equity-focused rating criteria were intended to elevate the ratings of priority group applicants because historically these groups have been underserved by MSAB. These criteria include:

- 3a. Applicant board and staff identify with underserved community/ies
- 3b. Significance of applicant’s work to underserved community/ies
- 3c. Applicant has a sustained connection to underserved community/ies

Figure 6 shows that these criteria worked as intended for BIPOC applicants (that is, BIPOC applicants received higher average ratings for these criteria as compared to non-priority applicants). For applicants led by or who serve people with disabilities, criteria 3a and 3b worked as intended (that is, these applicants received higher average ratings for these criteria as compared to non-priority applicants) and criterion 3c had no meaningful effect on average ratings. Lastly, for applicants from greater Minnesota, the criteria had no meaningful effect in average rating for these criteria (that is, there was no meaningful difference in average ratings for these criteria between greater Minnesota applicants and non-priority group applicants).

6. Equity-focused ratings for priority group applicants vs. non-priority group applicants

Criteria	BIPOC-led applicants	Applicants led by or who serve people with a disability	Greater Minnesota applicants
3a. Applicant identifies with underserved community/ies	Higher than non-priority	Higher than non-priority	No meaningful difference
3b. Significance of the applicant’s work to underserved community/ies	Higher than non-priority	Higher than non-priority	No meaningful difference
3c. Applicant has a sustained connection to underserved communities	Higher than non-priority	No meaningful difference	No meaningful difference

Note. “Higher than non-priority” refers to when a priority group’s average rating for these criteria was 10 or more percentage points *higher* than the non-priority group’s average rating. Similarly, “lower than non-priority” refers to when a priority group’s average rating for these criteria was 10 or more percentage points *lower* than the priority group’s average rating.

Overall, 330 organizations applied for a Creative Support for Organizations grant in the first two rounds of applications. Of those, 79 applicants identified as BIPOC, 3 applicants identified as having a disability, 14 applicants identified as providing services to people with disabilities, 132 applicants resided in greater Minnesota (outside of the Twin Cities metro), and 115 were non-priority group applicants. Please note that these figures total more than 330 because some applicants identified with more than one priority group. See Figure 7 for a breakdown of applicant type.

7. Percentage breakdown of applicant type

Applicant type	% of total applicants
BIPOC-led applicants	24%
Applicants led by people with disabilities	<1%
Applicants who serve people with disabilities	4%
Greater Minnesota applicants	40%
Non-priority group applicants	35%

Note. These percentages add to more than 100 percent because some applicants identified with multiple priority groups.

MSAB determined that all first and second round applicants (100%) could be funded with funds available through the Creative Support for Organizations program. Please see Figure 8 (on the next page) for the percentage of each applicant type that would have received funds if, for instance, MSAB could only fund 75% of applicants (similar to the Creative Support for Individuals program). In addition, we included a comparison of the total score funding allocation approaches to show how this allocation approach would have affected the applicant makeup for this grant program.

8. Percentage of successful applicants by applicant type for two funding allocation approaches

Applicant type	75% funding allocation	Total score funding allocation
BIPOC-led applicants	92% of applicants	85% of applicants
Applicants led by people with disabilities	100% of applicants	74% of applicants
Applicants who serve people with disabilities	100% of applicants	79% of applicants
Greater Minnesota applicants	76% of applicants	72% of applicants
Non-priority group applicants	60% of applicants	83% of applicants

Appendix

For the criteria with low reliability, see Figures A1 and A2 for some suggested updates that could improve reliability.

A1. Suggested sections from the Creative Support for Individuals review criteria that could benefit from updating

Criteria	High	Medium	Low
2a. Ideas can be accomplished with the resources available (time, money, people)	Ideas are clear and resources identified in the application (people, money) seem sufficient for making progress towards the outlined ideas within the grant year.	Ideas are somewhat vague, and/or some concerns whether resources identified in the application (people, money) are adequate for making progress towards the outlined ideas within the grant year.	Ideas were not sufficiently outlined, and/or progress could not be made towards ideas with the resources identified in the application (people, money).
<p><i>The highlighted yellow text may be what contributed to this criterion’s low reliability. There is not much distinction between “somewhat vague” and “not sufficiently outlined,” for example. Depending on reviewers’ backgrounds, experiences, or biases, they could likely rate the same application with a “medium” or “low” rating.</i></p>			
2b. Ideas informed by knowledge of audience and communities and/or the field	Applicant has a clear idea of who they, or their art, will connect with; ideas demonstrate strong knowledge of the field, or informed understanding of audiences/participants.	Applicant has considered who they or their art could connect with, but their ideas are vague. Ideas demonstrate basic knowledge of the field, or audiences/participants.	Applicant has not specifically identified who they or their art could connect with. It is not clear how ideas are informed by audiences/participants; lacking knowledge of the field.
<p><i>Similar to the previous criterion, the highlighted text may be what contributed to this criterion’s low reliability. For example, the distinction between “basic knowledge” and “lacking knowledge” is unclear. Depending on reviewers’ backgrounds, experiences, or biases, they could likely rate the same application with a “medium” or “low” rating.</i></p>			
3c. Sustained connection	Applicant demonstrates frequent and sustained connections with one or more underserved communities, through intentional or lived experience.	Applicant demonstrates connections with one or more underserved communities, but connections are not frequent or sustained.	There is no evidence that applicant connects with underserved communities.
<p><i>The highlighted yellow text may be what contributed to this criterion’s low reliability. There is not much clarity regarding what is meant by “frequent” or “sustained.” For instance, is frequent every week, every month, or every year? Similarly, does “sustained connection” refer to applicants who have years-long relationships with underserved communities, or to applicants engaging underserved communities every time they offer programming? Depending on reviewers’ backgrounds, experiences, or biases, they could likely rate the same application with a “high” or “medium” rating.</i></p>			

A2. Suggested sections from the Creative Support for Organizations review criteria that could benefit from updating

Criteria	High	Medium	Low
2b. Ideas informed by/incorporate input from communities	Applicant knows with whom they will be connecting; relevant community input (built-in, formal, or informal) is actively incorporated.	Applicant has considered who they could connect with, but their ideas are vague, it is unclear if community input is relevant or incorporated.	Applicant has not identified who they could connect with. There is no evidence of community input.
<p><i>The highlighted text may be what contributed to this criterion’s low reliability. There is not much clarity or distinction between “unclear if community input is relevant or incorporated” and “no evidence of community input,” for example. Depending on reviewers’ backgrounds, experiences, or biases, they could likely rate the same application with a “medium” or “low” rating.</i></p>			
3a. Applicant board and staff	50% or more of applicant board and staff identify as members of an underserved community, AND/OR; organization staff, board, and/or artists are inclusive of and deeply connected to people with disabilities.	Some staff or board identify as members of underserved communities, but no evidence that leadership roles are filled by members of underserved communities. OR Members of underserved communities serve on a permanent paid advisory council/committee that has some influence in organization decisions.	There is no evidence that applicant board and staff identify as members of an underserved community.
<p><i>The highlighted text may be what contributed to this criterion’s low reliability. In particular, the highlighted text could be read as contradictory. The first part of this rating notes that “no evidence that leadership roles are filled by members of underserved communities” but the second part notes a specific type of leadership—serving on a “permanent paid advisory council/committee” that influences organizational decisions. This contradiction makes it unclear when to give a “medium” rating. Our suggested edit is to delete the second part of the “medium” rating.</i></p>			

A2. Suggested sections from the Creative Support for Organizations review criteria that could benefit from updating (continued)

Criteria	High	Medium	Low
3b. Significance to underserved communities	Applicant demonstrates how the majority of their programming : 1) presents a unique perspective from within one or more communities, 2) provides representation in an art form for one or more underserved communities, or 3) specifically creates arts opportunities for underserved communities who are reflected in their board and staff .	Some of applicant's programming, but not a majority , is dedicated to presenting perspectives of, representation for, or arts opportunities for underserved communities.	Underserved communities may participate, but no evidence that any programming is specifically dedicated to underserved communities.
<i>The highlighted yellow text may be what contributed to this criterion's low reliability. It is unclear if applicants were instructed to present their programming in this way (i.e., the majority of their programming vs. less than the majority), which could make it difficult for reviewers to reliably rate applicants using this criterion. Additionally, the third part of the "high" rating brings another facet into this criterion—leadership makeup. This likely made it difficult to rate applicants using this criterion as well.</i>			
3c. Sustained connection	Applicant demonstrates a history of frequent, intentional, and sustained connections with audiences/participants/artists from underserved communities.	Applicant demonstrates a history of intentional connections with audiences/participants/artists from one or more underserved communities, but connections are not frequent or sustained .	There is no evidence that applicant has a history of making intentional connections with underserved communities.
<i>The highlighted yellow text may be what contributed to this criterion's low reliability. There is not much clarity regarding what is meant by "frequent" or "sustained." For instance, is frequent every week, every month, or every year? Similarly, does "sustained connection" refer to applicants who have years-long relationships with underserved communities, or to applicants engaging underserved communities every time they offer programming? Depending on reviewers' backgrounds, experiences, or biases, they could likely rate the same application with a "medium" or "low" rating.</i>			