NREPP Review Criteria: Criteria for Rating Quality of Research[1]
Each reviewer independently evaluates the Quality of Research for an intervention's reported results using the following six criteria:

1. Reliability

Outcome measures should have acceptable reliability to be interpretable. "Acceptable" here means reliability at a level that is conventionally accepted by experts in the field.

> 0 = Absence of evidence of reliability or evidence that some relevant types of reliability (e.g., test-retest, inter-rater, inter-item) did not reach acceptable levels.
>
> 2 = All relevant types of reliability have been documented to be at acceptable levels in studies by the applicant.
>
> 4 = All relevant types of reliability have been documented to be at acceptable levels in studies by independent investigators.

2. Validity

Outcome measures should have acceptable validity to be interpretable. "Acceptable" here means validity at a level that is conventionally accepted by experts in the field.

> 0 = Absence of evidence of measure validity, or some evidence that the measure is not valid.
>
> 2 = Measure has face validity; absence of evidence that measure is not valid.
>
> 4 = Measure has one or more acceptable forms of criterion-related validity (correlation with appropriate, validated measures or objective criteria); OR, for objective measures of response, there are procedural checks to confirm data validity; absence of evidence that measure is not valid.

3. Intervention Fidelity

The "experimental" intervention implemented in a study should have fidelity to the intervention proposed by the applicant. Instruments that have tested acceptable psychometric properties (e.g., inter-rater reliability, validity as shown by positive association with outcomes) provide the highest level of evidence.

> 0 = Absence of evidence or only narrative evidence that the applicant or provider believes the intervention was implemented with acceptable fidelity.
>
> 2 = There is evidence of acceptable fidelity in the form of judgment(s) by experts, systematic collection of data (e.g., dosage, time spent in training, adherence to guidelines or a manual), or a fidelity measure with unspecified or unknown psychometric properties.

---

4 = There is evidence of acceptable fidelity from a tested fidelity instrument shown to have reliability and validity.

4. Missing Data and Attrition

Study results can be biased by participant attrition and other forms of missing data. Statistical methods as supported by theory and research can be employed to control for missing data and attrition that would bias results, but studies with no attrition or missing data needing adjustment provide the strongest evidence that results are not biased.

0 = Missing data and attrition were taken into account inadequately, OR there was too much to control for bias.

2 = Missing data and attrition were taken into account by simple estimates of data and observations, or by demonstrations of similarity between remaining participants and those lost to attrition.

4 = Missing data and attrition were taken into account by more sophisticated methods that model missing data, observations, or participants, OR there were no attrition or missing data needing adjustment.

5. Potential Confounding Variables

Often variables other than the intervention may account for the reported outcomes. The degree to which confounds are accounted for affects the strength of causal inference.

0 = Confounding variables or factors were as likely to account for the outcome(s) reported as were the hypothesized causes.

2 = One or more potential confounding variables or factors were not completely addressed, but the intervention appears more likely than these confounding factors to account for the outcome(s) reported.

4 = All known potential confounding variables appear to have been completely addressed in order to allow causal inference between the intervention and outcome(s) reported.

6. Appropriateness of Analysis

Appropriate analysis is necessary to make an inference that an intervention caused reported outcomes.

0 = Analyses were not appropriate for inferring relationships between intervention and outcome, OR sample size was inadequate.

2 = Some analyses may not have been appropriate for inferring relationships between intervention and outcome, OR sample size may have been inadequate.

4 = Analyses were appropriate for inferring relationships between intervention and outcome. Sample size and power were adequate.