

Appendix C.
Assessment of IA-Led Evaluation
Review Form

ASSESSMENT OF IA-LED IMPACT EVALUATION
REVIEW FORM

To develop the evaluation review form, we started by emulating the data abstraction form that the Center for Substance Abuse Prevention (CSEP) used in development of the National Registry of Evidence-based Programs and Practices (NREPP) database, a service of the Substance Abuse and Mental Health Services Administration (SAMHSA; <http://www.nrepp.samhsa.gov/>). Then we compared the data abstraction form against the Society for Prevention Research Standards of Evidence criteria to ensure that the review form captured all relevant evaluation components (<http://www.preventionresearch.org/StandardsofEvidencebook.pdf>).

We expect raters to complete this review form after reading Implementing Agencies' (IA) State SNAP Ed Annual Final Reports and information extracted from other data sources as indicated in the accompanying matrix. We plan to collect much of the data for this review from data abstractions of IAs' applications and evaluation reports. Other data will be obtained from in-depth interviews with the evaluation manager at each of the IA sites.

Implementing Agency: _____

Reviewer: _____ **Date:** _____

Rating scale

The evaluation component being rated...		
Not Acceptable	1	...is missing or so poorly described that its value to the evaluation cannot be determined.
	2	...is inappropriate, misunderstood, or misrepresented in such a way that it cannot contribute to an effective evaluation of the program. The actions or materials reported are not appropriate from the evaluation effort proposed.
	3	...shows a general understanding of its role in the evaluation. However, key details have been overlooked or not thoroughly reported. Needs moderate revision to be considered acceptable.
Acceptable	4	...is appropriate for the evaluation, technically correct, and is described well enough to show a general understanding of its role in the overall evaluation. Evidence shows that it will or has been implemented properly, but minor details may be missing or unclear.
	5	...is appropriate for the program being evaluated and is presented in a way that shows the evaluator has a clear understanding of its role in the evaluation.

A. Research Objectives and Hypotheses

Score:

- **Clarity of research questions/hypotheses the evaluation is addressing**
 - *Are the objectives stated in SMART terms (specific, measurable, achievable, realistic, time-bound)?*
 - *A clear theory of causal mechanisms should be stated.*
- **Alignment of evaluation goals and objectives with intervention activities**
 - *Do the objectives/hypotheses include endpoints that are behavioral, meaningful, and related to the program's theory of change?*

B. Viable Comparison Strategy

Score:

(Outcome Evaluation Research Design)

Note: under no circumstances should self-selection into treatment or control be viewed as an acceptable method for developing a comparison strategy.

- **Appropriateness of the control or comparison group**
 - *Are the members of the control/comparison groups likely to be similar to the members of the treatment group? Is the study an experimental (randomized) or a quasi-experimental (non-randomized) design? Does this strategy make sense in the context of the treatment program?*
- **Threats to the validity of the design**
 - *Have plausible threats to validity (i.e., factors that permit alternative explanations of program outcomes) been discussed?*
 - *The evaluator must be able to rule out other factors that could explain changes, such as competing programs, concurrent media campaigns, and the effects of maturation among evaluation participants.*
 - *Absent true randomization, there is additional onus on the program to identify and rule out alternative explanations of program effects.*

C. Sampling Size/Sampling Strategy

Score: _____

- **Sample size estimations**
 - *Should be supported by power analysis that indicates the sample is sufficient to detect statistically significant differences in outcomes between treatment and control/comparison groups.*
 - *The power analysis should be matched to the outcome evaluation design. It should be based on an anticipated program effect size that is empirically valid (i.e., drawn from published literature or pilot work).*
- **Method of selecting sample participants from the population.**
 - *Should specify what/who the sample is and how it was obtained. Should be detailed and provide a reasonable basis for generalization of program effects to the broader population of people 'like those' in the study.*
- **Recruitment plans.**
 - *Description of steps to be taken by project staff to increase the likelihood that members of the target population approached by the program will agree to participate in the program*
NOTE: no program will have 100% recruitment, but rates below 70% - 80% should be closely examined for justification.

D. Outcome Measures

Score: _____

- **Quality of the data collection instruments (surveys, interviews)**
 - *Information on reliability (internal consistency (alpha), test-retest reliability, and/or reliability across raters) and construct validity of measures should be provided.*
 - *When possible, the use of scales is preferable to single item measures.*
- **Alignment of evaluation measures with the intervention activities.**
 - *Outcome measures assess actual behavior change.*
 - *Outcome measures should map onto research objectives/hypotheses*
 - *Higher scores should be considered for measures that include intermediate factors in the behavior change process.*

E. Data Collection

Score: _____

- **Overview of data collection schedule**
 - *Timing of data collection should align with program activities*
 - *Should be realistic and achievable*
- **Rigor of the data collection process**
 - *Data collection for the intervention and comparison group participants should be similar. Any differences should be noted and justified.*
 - *Participant data should be anonymous (no names linked to data) or confidential (names linked to data are kept private).*
 - *Should include description of data management and data security measures*
 - *Describe longitudinal tracking procedures*
- **Quality of the data collection process**
 - *Evidence of thorough training of data collectors*
 - *High scores should be given for data collection procedures that are least likely to introduce bias or promote non-response.*

F. Data Analysis

Score: _____

Note: Descriptive statistics are not sufficient to show program effects!

- **Sample characteristics and baseline comparability**
 - *Tables showing demographic information and number of participants in the intervention and comparison groups*
 - *Statistical tests assessing baseline comparability across treatment conditions*
- **Statistical methods used to assess the program impacts**
 - *Multivariate statistics should be used to assess program effects*
 - *Statistical approach should be matched to the characteristics of the research design and the data being collected*
 - *Highest scores should be given to programs that include meditation analyses*
- **Additional Statistical Procedures and Analyses**
 - *Analyses/Methods for handling attrition bias are proposed/conducted properly*
 - *Procedures for accounting for missing data are proposed/conducted properly*
 - *Subgroup analyses proposed/presented for primary outcomes*
Potential indicators for specifying sub-groups include demographic and socioeconomic variables.

G. Attrition (loss of participants)

Score: _____

- **Attrition is program drop out.** *It is the differences between the number of participants completing baseline survey and the number completing the post-intervention and follow-up survey(s). Modest attrition should be anticipated in the design. Lowest scores given for extraordinary attrition rates.*

H. Missing Data (incomplete survey/items)

Score: _____

- **Missing data is survey non-response.** *It represents the absence of, or gaps in, information from participants who remain involved in the evaluation. Lowest scores given for a large amount of missing data.*