



Continuum of Evidence <sup>1 2</sup>				
Placement	Effective <sup>3</sup>	Promising	Unclear <sup>4</sup>	Ineffective
Criteria				
Significant Effect	Rigorous statistical evidence of a change in a highly desired outcome that was considered significant, with no negative effects found.	Rigorous statistical evidence of a change in a highly desired outcome that was considered significant, with no negative effects found.	Effects are unclear due to mixed results or no evidence.	An appropriate evaluation has failed to demonstrate a significant effect, or has negative effects.
Sustained Effect	Effect(s) lasting ≥ two years from the beginning of the program, or ≥ one year from program completion.	Effect(s) lasting ≥ one year from the beginning of the program, or ≥ 6 months from program completion. <i>Noted considerations may be given for programs that have not had sufficient time to demonstrate long-term effects.</i>	Sustainability not assessed or established.	Program effects not sustained.
Successful External Replication	Program was found effective in at least one other study that matches the original evaluation study design, and conducted by an implementation team that was *independent of the program developer.	No evidence of external replication, or limited replication criteria (i.e., lacking significant/sustained effect, inadequate study design, etc.).	No evidence of external replication.	No evidence of successful external replication.
Study Design	Randomized controlled design Well-matched quasi-experimental design	At least a quasi-experimental design	May use a quasi-experimental, pre-post-test design, or purely descriptive	Experimental or quasi-experimental design
Additional Criteria Regarding Study Execution	Meets all 4 additional criteria (see pages 2-3).	Meets 2 or 3 of the additional criteria (see pages 2-3).	Meets 0 or 1 additional criteria (see pages 2-3).	Meets all 4 additional criteria (see pages 2-3).

<sup>1</sup>Adapted from two sources: (1) Blueprints for Violence Prevention (<http://www.colorado.edu/cspv/blueprints/>) and (2) OJP What Works Repository (<http://www.ncjrs.gov/pdffiles1/nij/220889.pdf>).

<sup>2</sup>This material is based upon work supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, under Agreement No. 2010-488869-20781 as part of the USDA National Institute of Food and Agriculture - Department of Defense Military Community and Family Policy Partnership.

<sup>3</sup>The Effective category has two subcategories: (1) Effective – RCT, and (2) Effective – Quasi-experimental

<sup>4</sup>The Unclear category has three subcategories: (1) Unclear (Ø) with no evaluations or mixed results, (2) Unclear (+) with potentially promising features, and (3) Unclear (-) with potentially ineffective features.



## Effective – RCT Program Category on the Continuum of Evidence:

### Description:

A program with an experimental research design (RCT) that:

- demonstrates a significant and sustained effect;
- has at least one replication involving an external implementation team at a different site (from the original successful study); and
- meets all of the additional criteria listed below regarding study design and execution.

### Criteria:

A. Significant Effect— rigorous statistical evidence of a change in a highly desired outcome (i.e., the behavioral, psychological, knowledge, or attitude change that is the stated goal of the program) that was considered significant. In addition to main effects of the intervention, effects of implementation level and dosage analyses may be considered. Two-tailed tests of significance are preferable to one-tailed test. They are more conservative and test for effects in both directions.

B. Sustained Effect— effect(s) lasting  $\geq$  two years from the beginning of the program, or  $\geq$  one year from program completion.

C. At Least One Successful External Replication— Program was found to have sustained effects in at least one additional randomized controlled trial (RCT) conducted by an implementation team that was \*independent of the program developer.

- Free of any connection with the developer (i.e., initial implementation team, previous or current student, etc.)
- Adaptations: only in the conservative direction (ex. significant effects found with 9 vs. 12 sessions, and 75% of major outcomes found effective).
- No differences in diagnostic category of study sample

D. Randomized controlled design

E. All 4 of the following additional criteria have been met:

- Representative Sample – The study sample accurately represents the population that the program purportedly targets. If the study does not fully represent the target audience, exceptions can be made, but must be clearly described in the Evidence section of the fact sheet. For example, if a program states that it is for all elementary school children, but it is only tested on 4<sup>th</sup> and 5<sup>th</sup> graders, you may place it as Promising for 4<sup>th</sup> and 5<sup>th</sup> graders, but must also mention that it is Unclear for Kindergarteners through 3<sup>rd</sup> graders. Replication only counts if it is on similar samples (e.g., same age and same diagnosis).
- Modest Attrition – Modest attrition is  $\leq$  20%. However, attrition may vary by length of study. Moreover, there is always a concern if attrition  $>$  50%. Look to see if they compared those who left the study to those who stayed. A rough guide: attrition at immediate post test  $\leq$  10%; attrition at 6 month follow up  $\leq$  15%; attrition at 1 year follow up  $\leq$  20%; attrition at 2 year follow up  $\leq$  30%. Attrition may compromise the randomization process if the attrition itself is nonrandom. High attrition makes it more difficult to be confident that effects are due to treatment and not something else. If attrition exceeds the above recommendations, but sample size and attrition are reported at each follow-up, and tests show that attrition is not differential, then it may be given a “yes” as opposed to a “no”.
- Practical Significance – An explanation of practical, vs. statistical, significance is provided. This may be related to effect sizes, movement from above a clinical cutoff to below a clinical cutoff or vice versa depending on the outcome, or discussion of actual change in behavior (is reduction in alcohol use .025 drinks per day or 1 drink per day).
- Adequate Outcome Measurement - Reliable and valid assessments are used to measure outcomes. In addition, outcome measurement matches stated goals of the program (e.g., if stated goals of the program are to reduce alcohol use, outcomes must measure reduction in alcohol use, and not just attitudes about alcohol use). Also, analyses should be conducted at the same level as the randomization.

### Anecdotal Data:

Anecdotal data is a practical source of data for professionals implementing programs because it can provide useful information about the participant’s responsiveness to the program as well as being an important part of telling a program’s story. Examples of anecdotal data include participants’ comments about the program and participants’ satisfaction (e.g., I like the program and it was fun). Anecdotal data does not provide evidence on the overall effectiveness of a program, only a snapshot of select individuals’ experiences with a program. Therefore, anecdotal data does not impact the placement of a program on the Clearinghouse’s Continuum of Evidence.



## Effective – Quasi-experimental Program Category on the Continuum of Evidence:

### Description:

A program with a quasi-experimental research design that uses well-matched or statistically controlled comparison groups that:

- demonstrates a significant and sustained effect;
- has at least one replication involving an external implementation team at a different site (from the original successful study); and
- meets all of the additional criteria listed below regarding study design and execution.

### Criteria:

A. Significant Effect— rigorous statistical evidence of a change in a highly desired outcome (i.e., the behavioral, psychological, knowledge, or attitude change that is the stated goal of the program) that was considered significant. In addition to main effects of the intervention, effects of implementation level and dosage analyses may be considered. Two-tailed tests of significance are preferable to one-tailed test. They are more conservative and test for effects in both directions.

B. Sustained Effect— effect(s) lasting  $\geq$  two years from the beginning of the program, or  $\geq$  one year from program completion.

C. At Least One Successful External Replication— Program was found to have sustained effects in at least one additional quasi-experimental design study conducted by an implementation team that was \*independent of the program developer.

- Free of any connection with the developer (i.e., initial implementation team, previous or current student, etc.)
- Adaptations: only in the conservative direction (ex. significant effects found with 9 vs. 12 sessions, and 75% of major outcomes found effective).
- No differences in diagnostic category of study sample

D. Quasi-experimental research design

E. All 4 of the following additional criteria have been met:

- Representative Sample – The study sample accurately represents the population that the program purportedly targets. If the study does not fully represent the target audience, exceptions can be made, but must be clearly described in the Evidence section of the fact sheet. For example, if a program states that it is for all elementary school children, but it is only tested on 4<sup>th</sup> and 5<sup>th</sup> graders, you may place it as Promising for 4<sup>th</sup> and 5<sup>th</sup> graders, but must also mention that it is Unclear for Kindergarteners through 3<sup>rd</sup> graders. Replication only counts if it is on similar samples (e.g., same age and same diagnosis).
- Modest Attrition – Modest attrition is  $\leq$  20%. However, attrition may vary by length of study. Moreover, there is always a concern if attrition  $>$  50%. Look to see if they compared those who left the study to those who stayed. A rough guide: attrition at immediate post test  $\leq$  10%; attrition at 6 month follow up  $\leq$  15%; attrition at 1 year follow up  $\leq$  20%; attrition at 2 year follow up  $\leq$  30%. Attrition may compromise the randomization process if the attrition itself is nonrandom. High attrition makes it more difficult to be confident that effects are due to treatment and not something else. If attrition exceeds the above recommendations, but sample size and attrition are reported at each follow-up, and tests show that attrition is not differential, then it may be given a “yes” as opposed to a “no”.
- Practical Significance – An explanation of practical, vs. statistical, significance is provided. This may be related to effect sizes, movement from above a clinical cutoff to below a clinical cutoff or vice versa depending on the outcome, or discussion of actual change in behavior (is reduction in alcohol use .025 drinks per day or 1 drink per day).
- Adequate Outcome Measurement - Reliable and valid assessments are used to measure outcomes. In addition, outcome measurement matches stated goals of the program (e.g., if stated goals of the program are to reduce alcohol use, outcomes must measure reduction in alcohol use, and not just attitudes about alcohol use). Also, analyses should be conducted at the same level as the randomization.

### Anecdotal Data:

Anecdotal data is a practical source of data for professionals implementing programs because it can provide useful information about the participant’s responsiveness to the program as well as being an important part of telling a program’s story. Examples of anecdotal data include participants’ comments about the program and participants’ satisfaction (e.g., I like the program and it was fun). Anecdotal data does not provide evidence on the overall effectiveness of a program, only a snapshot of select individuals’ experiences with a program. Therefore, anecdotal data does not impact the placement of a program on the Clearinghouse’s Continuum of Evidence.



## Promising Program Category on the Continuum of Evidence:

### Description:

A program with at least a quasi- experimental design without a successful replication, or a program with a prospective, quasi-experimental research design using well-matched comparison groups that have significant and sustained effects, and meets at least 2 of the additional criteria regarding study design and execution.

### Criteria:

- A. Significant Effect— statistical evidence of a change in a highly desired outcome (i.e., the behavioral, psychological, knowledge, or attitude change that is the stated goal of the program) that was considered significant. In addition to main effects of the intervention, effects of implementation level and dosage analyses may be considered. Two-tailed tests of significance are preferable to one-tailed test. They are more conservative and test for effects in both directions.
- B. Sustained Effect — effect(s) lasting  $\geq$  one year from the beginning of the program, or  $\geq$  6 months from program completion. *Noted considerations may be given for programs that have not had sufficient time to demonstrate long-term effects.*
- C. Successful External Replication — no evidence of external replication, or limited replication criteria (i.e., lacking significant/ sustained effect, inadequate study design, etc.).
- D. Experimental or quasi-experimental study design
- E. Meets 2 or 3 of the following additional criteria: (see definitions above)
  - Representative sample;
  - Modest attrition;
  - Practical significance;
  - Adequate outcome measurement.

### Anecdotal Data:

Anecdotal data is a practical source of data for professionals implementing programs because it can provide useful information about the participant's responsiveness to the program as well as being an important part of telling a program's story. Examples of anecdotal data include participants' comments about the program and participants' satisfaction (e.g., I like the program and it was fun). Anecdotal data does not provide evidence on the overall effectiveness of a program, only a snapshot of select individuals' experiences with a program. Therefore, anecdotal data does not impact the placement of a program on the Clearinghouse's Continuum of Evidence.



## Unclear Program Category on the Continuum of Evidence:

**Description:** An unclear program may be described in one or more of the following ways.

- A program with (1) a quasi-experimental design that lacks sufficient methodological rigor; (2) a pre-post-test design without a comparison group; or (3) purely descriptive or qualitative evaluation (e.g. case studies); and
- Meets 0 or 1 of the following additional criteria: (see definitions above)
  - Representative sample;
  - Modest attrition;
  - Practical significance;
  - Adequate outcome measurement.

Program placers will now have three options for the Unclear Placement. Please choose one based on the following criteria and make a note of WHY you chose this placement.

(1) **Unclear (Ø) with no evaluations or mixed results**– if no peer reviewed evaluations have been identified; or if there are mixed results or other issues that do not suggest that it is clearly leaning towards promising or ineffective.

- Example: There are no peer-reviewed evaluations of the program. A program claims to have an effect on two outcomes, but fails to find significant positive results for one of the two outcomes (especially if the two outcomes should be correlated). One study found positive results, but another equally well designed study failed to find positive results. A pre-post evaluation with a small sample size has failed to find effects.

(2) **Unclear (+) with potentially promising features** – if there are promising or effective aspects, but due to our criteria falls into the Unclear placement.

- Example: An RCT has shown positive effects, but it only goes out 3 months post intervention.

(3) **Unclear (–) with potentially ineffective features** – if, due to our criteria, it falls into the Unclear placement, but there is reason to believe that it may be ineffective.

- Example: A pre-post evaluation fails to find significant effects. A quasi-experimental study fails to find significant effects but there are questions about the power related to the sample size.

### Anecdotal Data:

Anecdotal data is a practical source of data for professionals implementing programs because it can provide useful information about the participant's responsiveness to the program as well as being an important part of telling a program's story. Examples of anecdotal data include participants' comments about the program and participants' satisfaction. Anecdotal data does not provide evidence on the overall effectiveness of a program, only a snapshot of select individuals' experiences with a program. Therefore, anecdotal data does not impact the placement of a program on the Clearinghouse's Continuum of Evidence. Indeed, anecdotal data may suggest positive responses from participants for a program in this category; however, the objective evidence does not support a program's effectiveness.



## Ineffective Program Category on the Continuum of Evidence:

**Description:** A program with a strong research design that in an initial study and at least one *appropriate* additional evaluation has failed to demonstrate a significant effect, or has demonstrated significant negative effects.

For a program to be placed as INEFFECTIVE, the criteria would need to be as rigorous as an effective or promising program.

### **Anecdotal Data:**

Anecdotal data is a practical source of data for professionals implementing programs because it can provide useful information about the participant's responsiveness to the program as well as being an important part of telling a program's story. Examples of anecdotal data include participants' comments about the program and participants' satisfaction. Anecdotal data does not provide evidence on the overall effectiveness of a program, only a snapshot of select individuals' experiences with a program. Therefore, anecdotal data does not impact the placement of a program on the Clearinghouse's Continuum of Evidence. Indeed, anecdotal data may suggest positive responses from participants for a program in this category; however, the objective evidence does not support a program's effectiveness.