ASSESSING LEVELS OF EVIDENCE

Overview

For grantmaking institutions applying to the Social Innovation Fund (SIF) with a pre-identified intervention, the evidence level of that intervention will be assessed during the application review. This document explains how that assessment will be conducted and includes the rubric that will be used.

The process of determining the level of evidence for an intervention can be complicated and often requires technical knowledge of research and evaluation design and methods. The rubric introduced in this document provides a framework for assessing the existing body of evidence based on past research and evaluation studies.

What Does "Attaining a Level of Evidence" Mean?

As described in the SIF NOFA, all SIF-funded interventions require at least a Preliminary level of evidence upon entering the SIF. This level of evidence will be demonstrated by studies conducted prior to applying for SIF funding.

- To attain the **Preliminary** level of evidence required for SIF funding, an intervention must, at a minimum, have a study that has "yielded promising results for either the program or a similar program." Specifically, the intervention must have at least some outcome information such as *pre- and post-tests* without a comparison group, or *post-test comparison between program and comparison groups*.
- To attain a **Moderate** level of evidence, an intervention needs to have evidence "from studies whose designs can support causal conclusions (i.e., studies with high internal validity[1]), but have limited generalizability (i.e., moderate external validity[2]), or studies with high external validity, but moderate internal validity."[3]

  Studies with high internal validity will likely use *Quasi-experimental Designs* (QED) (such as a matched comparison group or a comparative interrupted time series design) or *Randomized Controlled Trials* (RCT) also known as Experimental Designs. At least one study with high internal or external validity is typically needed to attain a **Moderate** level of evidence.
- To attain a **Strong** level of evidence, an intervention should have designs that "can support causal conclusions (i.e., studies with high internal validity), and studies that in total include enough of the range of participants and settings to support scaling up to the state, regional, or national level (i.e., studies with high external validity)." Interventions that enter the SIF with a **Strong** level of evidence

[1] Internal validity for a study is the extent to which the observed difference in the average group outcomes (usually program participants versus control or comparison group members) can be causally attributed to the intervention or program.
[2] External validity for a study is the extent to which evaluation results are applicable to groups other than those in the research.
[3] Moderate internal validity could come from a study having a comparison group formed without statistical matching techniques, statistical matching techniques that resulted in lower than desirable pre-test group equivalence, or an interrupted time series design without a comparison group

would have conducted either *one large, multisite RCT or QED study* or *several smaller RCT or QED studies either in different locations or with different populations*.

Assessing Your Incoming Level of Evidence

Although there are several factors to consider when assessing an intervention's level of evidence, this rubric focuses on two important sets of factors:

1) Similarity of the intervention under consideration to the previously studied intervention(s) in terms of where and how they were implemented and
2) Type of study or studies previously conducted.

Similarity of the Intervention Under Consideration to Previously Studied Interventions

Identify how the previously studied interventions relate to the intervention you are considering in the following ways:

- Was the intervention implemented by your organization or a different one?
- How closely matched is the previously studied intervention to the proposed intervention?
- In other words, was the studied intervention identical or very similar to the proposed intervention in terms of content, delivery or target population, or was it substantially modified, adapted, or combined with other interventions?

Type of Study Conducted

Identify which types of research or evaluation designs were used in prior studies of the proposed intervention. Only consider studies that yielded positive results.

- What types of studies showed positive results - rather than null or negative results - for the outcomes targeted by the applicant program?
- For example, are there positive results from studies that have used designs such as pre- and post-tests with a single group? Are there studies that used a matched comparison group? Was there a randomized controlled trial?

Issues to Consider

Additional issues to consider when assessing incoming evidence:

- Adapting an intervention or combining multiple interventions may lower the assessed evidence level.
- Even with the same study design (e.g., a single site RCT), an intervention using evidence from studies of a *similar* intervention may have a lower assessed level of evidence than an intervention using studies from the *identical* intervention.

- Unless an intervention is being intentionally replicated with fidelity, studies for the same intervention conducted by a different program or organization may also offer lower levels of evidence than studies conducted by the proposing subgrantee.
- A study or studies conducted in a different organizational context than the one being proposed does not likely have sufficient evidence to be considered preliminary under SIF standards. This is due to the fact that the preliminary evaluations (i.e., single group pre-post- tests) do not have sufficient internal validity to show that the program "causes" the outcome. For studies that only offer pre-post testing, it is possible that something in the program context other than the intervention (e.g., how participants are selected) may be causing the changes seen by that program.

Using the Evidence Level Rubric

To use the Evidence Level Review Rubric on the next page and find an intervention's incoming level of evidence:

- Review each previously conducted study. Identify those that generally show positive, rather than null or negative, results for the outcomes targeted by the applicant's program.
- Determine the connection of the proposed intervention to the studied intervention and use the labels in the top row, "Similarity to Proposed Intervention," to select the column that best represents how that study relates to the proposed intervention.
- Put a check in the box(es) of the column selected in step 2 above, where it intersects with the row for the design type used in the study from the choices in the left hand column "Study Design."
- Review each study and check each relevant box.
- After following this procedure for each study, the highest-ranked checked box (e.g., Preliminary, Moderate, Strong) is the level of evidence for the proposed intervention.

INCOMING EVIDENCE LEVEL RUBRIC

Intervention: **Highest Ranked Checked Box:**

Columns: How matched – by organization and similarity - is the previously studied intervention to the proposed Intervention? Was it done by…

Rows: What type of design was used for studies where the results on relevant outcomes were positive?

	A different organization doing a similar, but not identical intervention?	A different organization doing an identical intervention? (Proposed intervention will be replicated with fidelity)	The same organization doing a combination of interventions that includes the one studied?	The same organization doing an intervention that is similar, but not identical to the studied intervention?	The same organization, doing exactly the same intervention?
None or not known	Not yet preliminary ☐	Not yet preliminary ☐	Not yet preliminary ☐	Not yet preliminary ☐	Not yet preliminary ☐
Implementation only	Not yet preliminary ☐	Not yet preliminary ☐	Not yet preliminary ☐	Not yet preliminary ☐	Not yet preliminary ☐
Pre-post testing	Not yet preliminary ☐	Not yet preliminary ☐	Not yet preliminary ☐	Not yet preliminary/ Preliminary (Depends on extent of similarity) ☐	Preliminary ☐
Pre-post or post only with non-matched comparison group, or interrupted time series with no comparison group	Not yet preliminary ☐	Not yet preliminary ☐	Preliminary ☐	Preliminary ☐	Preliminary ☐
Single site, well designed and implemented QED or RCT	Preliminary ☐	Preliminary ☐	Preliminary ☐	Preliminary ☐	Moderate ☐
Two or three well designed and well implemented single site RCTs or QEDs	Preliminary ☐	Moderate ☐	Preliminary ☐	Preliminary/Moderate (Depends on extent of modification) ☐	Moderate ☐
National/large scale multi-site well designed and well implemented QED or RCT, or multiple (three or more) well designed and well implemented QEDs or RCTs in different locations	Preliminary ☐	Strong ☐	Preliminary ☐	Preliminary/Moderate (Depends on extent of modification) ☐	Strong ☐

**(To be designated Strong+ and to be exempted from the requirement to attain moderate evidence with its SIF evaluation, the program would need an extensive, multi-site history of RCT's/QED's with the population in question.)*