

Best Practices in Demonstrating Evidence



Informational Webinar

A presentation for AmeriCorps grant applicants

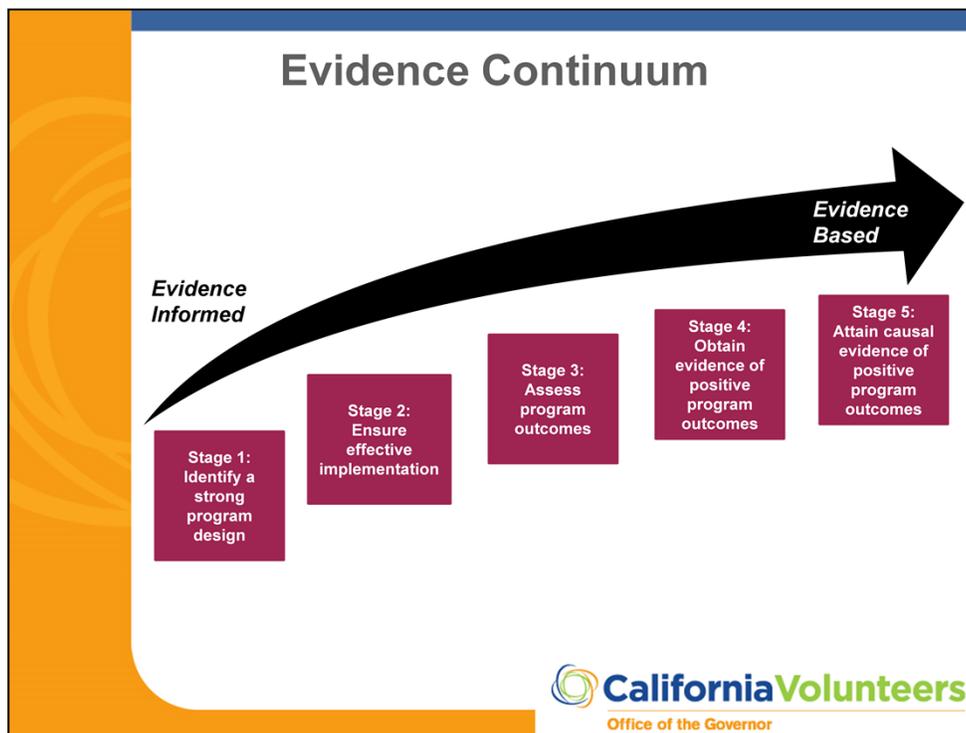


Overview

- Evidence in national service
- Overview of basic evaluation concepts
- Overview of evidence tiers
- Q&A

Why is Evidence Important?

- To test whether programs are effective, and what makes them effective
- To ensure that public dollars are invested wisely
- To inform continuous improvement of programs
 - Change what isn't working
 - Do more of what is working



The AmeriCorps Evidence Continuum

- We recognize that it takes time to build evidence and that the strength of a program's evidence varies according to where the program is in its life cycle
- The process of building evidence begins with a strong theory of change and logic model
- Over time, programs collect data about implementation and outcomes and eventually engage in more rigorous types of evaluation to determine if there is a cause and effect relationship between what AmeriCorps members are doing and the program's outcomes
- Programs should know where they are on the continuum and have a plan for building evidence over time

Evidence in the 2016 RFA

- Evidence section is worth 10 points
- Points awarded based on strength and quality of evidence (evidence tiers)
- Applicants should determine the highest evidence tier for which they are eligible and describe their evidence clearly, completely, and accurately



We expect you to incorporate evidence as appropriate throughout the application. The Evidence Base section is where you will describe the body of evidence for the effectiveness of your proposed intervention and describe up to two studies for moderate or strong evidence.

2015 CNCS Staff Reviewed vs. Funded

Level of Evidence	Staff Review	Funded	Percent (Funded/Staff Review)
Strong	8	8	100%
Moderate	14	14	100%
Preliminary	97	59	61%
Pre-Preliminary	63	22	35%
No Evidence	38	11	29%
Total	220	114	52%



Just to give you an idea of evidence levels in last year's federal competition

- All applications with moderate or strong evidence that made it to staff review were funded.
- Over half of all applications reviewed with preliminary evidence were funded.
- Applications with pre-preliminary evidence and no evidence did not fare quite as well; however, applicants at these levels may also have other capacity issues or may be new to AmeriCorps. It is still possible for pre-preliminary and preliminary programs to get funded if their program design and application are of high quality.

2015 Grantee Levels of Evidence

Level of Evidence	Percent
Strong	7%
Moderate	12%
Preliminary	52%
Pre-Preliminary	19%
No Evidence	10%
Total	100%



Most of our field is in the early stages of evidence building. In 2015, 52% of successful grantees had preliminary evidence. 29% had pre-preliminary or no evidence.

Performance Measurement and Evaluation

Performance Measurement	Program Evaluation
<ul style="list-style-type: none">• Ongoing monitoring and reporting of program accomplishments and progress• Explains what level of performance is achieved by the program	<ul style="list-style-type: none">• In-depth research activity conducted periodically or on an ad-hoc basis• Answers questions or tests hypotheses about program processes and/or outcomes• Used to assess whether or not a program works as expected and why (e.g., did the program cause the observed changes?)

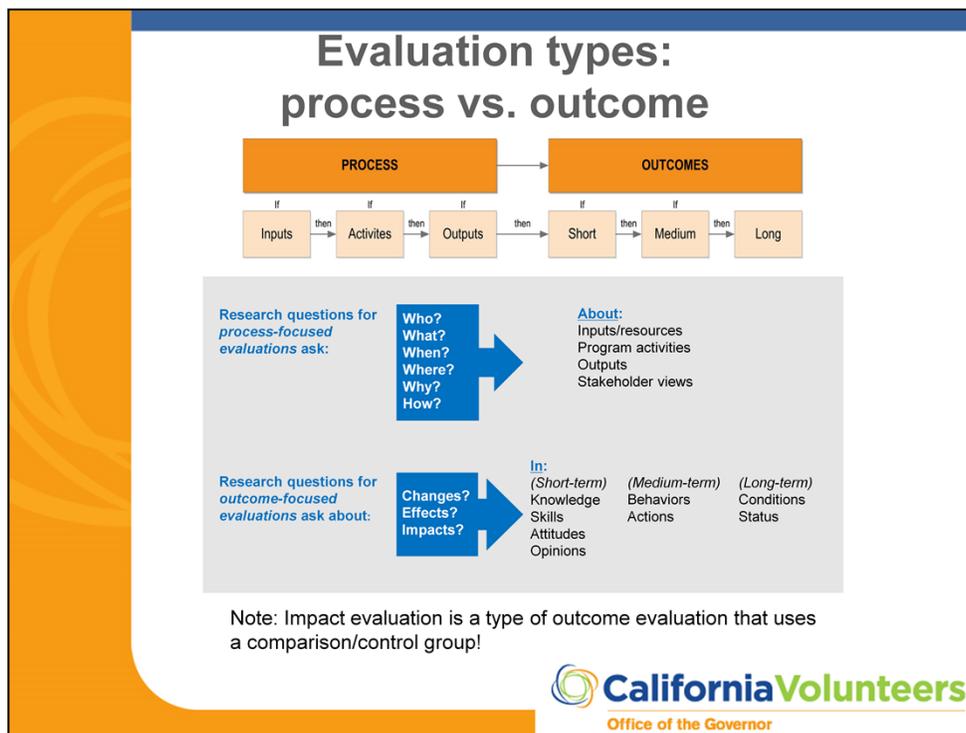


Performance measurement monitors a few vital signs related to program performance objectives, outputs, and/or outcomes.

Evaluation, on the other hand, comprehensively examines programs using systematic, objective, and unbiased procedures in accordance with social science research methods and research designs.

Performance measurement is used to gauge program delivery and participant satisfaction for continuous improvement, enhanced decision making, and program development. Evaluation is used to assess program effectiveness and determine whether the program is responsible for changes found.

Program staff typically manage the ongoing performance measurement processes; an experienced researcher conducts the evaluation periodically.



Facilitator notes: The results of a process evaluation are most often used to change or improve the program. A process evaluation can be used to document what a program is doing and to what extent and how consistently the program demonstrates fidelity to the program’s logic model. An outcome or impact evaluation can be used to determine the results or effects of a program. These types of evaluations generally measure changes in program beneficiaries’ knowledge, attitudes, behaviors, and/or conditions thought to result from the program.

Process evaluations address exploratory questions about program operations, processes, and procedures, namely the who, what, when, where, why, and how of program activities and program outputs. On the other hand, outcome evaluations test hypotheses about a program’s changes, effects, or impacts on short-, medium-, or long-term outcomes, ultimately assessing program effectiveness.

Impact evaluation is a type of outcome evaluation that utilizes a comparison or control group. It is the only type of evaluation that can attribute causation to the program.

Evaluation Designs

Evaluation Study Designs	Comparison	Ability to make statements about causal attribution
Experimental Design Studies	Randomly Assigned Groups	↑
Quasi-Experimental Design Studies	Statistically Matched Groups	
Non-Experimental Design Studies	Not Statistically Matched Groups or Group Compared to Itself	

-Here we show the different evaluation study designs have the ability to make statements about causal attribution.

-Looking at the bottom row first – non experimental design studies do not have any kind of comparison or control group. These studies cannot make causal statements.

-Moving one row up to quasi- experimental designs – these studies use statistically matched control groups and can be used to establish causality if the matching is high-quality and statistically valid.

-Finally, experimental design studies, or randomized control trials, use randomly assigned groups. These are the strongest research designs and have the ability to make causal statements about the intervention.

Evidence Tiers: No evidence and pre-preliminary

- **No evidence**
 - Applicant has not systematically collected any qualitative or quantitative data on their program
- **Pre-preliminary**
 - Applicant has collected systematic and accurate data to test or track one or more components of its logic model (ex: community need, outputs, participant outcomes) OR
 - Applicant has conducted a process evaluation assessing implementation of one or more interventions depicted in the logic model
 - The data collection process and results are described fully
 - The applicant explains the link between data collection and the relevant component(s) of its logic model

No Evidence - Example

Narrative: Applicant A's mentoring program incorporates the Elements of Effective Practice for mentoring, a set of evidence-based standards for mentoring programs. The program is modeled closely on Famous Mentoring Program's successful approach. A 2013 randomized control trial found Famous Mentoring Program to be effective.

Additional Documents: The applicant submitted a copy of Famous Mentoring Program's successful approach.



Incorporating best practices from your field or from a similar program, even if those best practices are evidence-based, does not constitute evidence for your program. Programs that incorporate evidence from other programs into their program design are considered evidence-informed. An evidence informed program uses the best available knowledge, research, and evaluation to guide program design and implementation, but does not have scientific research or rigorous evaluation of the program itself and is not replicating an evidence-based program. While evidence informed programs do not receive points in the evidence section, reviewers will also consider how evidence informs the program's theory of change when assessing the theory of change/logic model criteria.

In this case, the applicant says it is modeled closely on another program that conducted a RCT; however, the applicant does not describe how it is replicating this other program with fidelity. Simply being similar to, or incorporating some practices from, an evidence-based model is not sufficient. It is also not sufficient to say that you will replicate an evidence-based program without fully explaining how you will replicate it. Because the program is not replicating the Famous Mentoring Program with fidelity, they are not eligible to submit that program's study for consideration.

Pre-Preliminary Example

In Applicant B's last full year of operations, we provided tutoring services to 500 students (ED1, target was 475). 452 students completed the required dosage of 30 minutes, twice a week for 6 months (ED2, target was 450). Of these 450 students, 350 met our improvement benchmark for Performance Measure ED5 (target was 300)— a gain of at least one grade level on the Famous Standardized Literacy Assessment. This standardized test measures reading comprehension and has demonstrated validity and reliability for the population of second and third graders served by our program. It is administered as a pre-test when students enter the program and again at the end of the program. This gain is significant given that most students begin the program 2-3 grade levels behind and would not have been expected to make a year's improvement in six months without significant support from tutors. Improving academic engagement remains a primary focus of our program in 2016, and we have included these performance measures in our logic model.



This example is strong because it:

- States the performance measure outputs and outcomes (and references numbers of National Performance Measures, if applicable)
- Provides numerical values for targets and actuals
- States the dosage that children in the program received – the reviewer can look at the applicant's logic model to see if proposed dosage is the same
- Explains why the amount of improvement was significant
- Describes when data was collected
- Describes the instrument that was used to collect data

The reason this does not qualify as preliminary is because there was not an outcome study that went beyond regular performance measurement.

Evidence Tiers: Preliminary (1)

- **Preliminary – Option 1 – Outcome study of own program**
 - Applicant has conducted at least one outcome study of its own intervention, either pre and post-test without a comparison group or post-test only with a comparison group
 - The outcome study includes data beyond that which is collected as part of routine performance measurement
 - The applicant provides a detailed description of the outcome study data
 - The description explains whether the outcome study was conducted by the applicant organization or by an entity external to the applicant
 - The outcome study yielded promising results for the proposed intervention

Preliminary Evidence – Example 1 – Outcome study of own program

Applicant C is a small program focused on helping homeless individuals gain knowledge of responsible tenant practices and other housing support resources, and ultimately find and maintain affordable housing. In our last complete program year, 250 homeless individuals received housing services (O5, target = 200) and 200 of these individuals were transitioned into safe, affordable housing (O10, target = 175). Since 2011 we have sent a follow-up survey nine months after an individual was transitioned into housing to determine whether they remained housed. We analyzed this survey data for our 2014 outcome evaluation and found that 95% of individuals responding to the survey remained in affordable housing, a rate much higher than the national average of 80% for the population we serve.



What's good about this example:

- Discusses their outcome evaluation, which is the best evidence they have
- Also discusses their PM data, so we have the full picture of the program's evidence
- Provides comparative statistics (national average of 80%) to put results of outcome evaluation in context
- The nine-month follow up survey is appropriate for outcome evaluation as it goes beyond what is collected for performance measurement outcome

What could be better:

- Does not say if the outcome study was conducted by an internal or external evaluator although "we" implies internal.
- Does not discuss survey response rate and potential non-response bias.

Evidence Tiers: Preliminary (2)

- **Preliminary – Option 2 – Replication with fidelity**
 - Applicant is proposing to replicate an evidence-based program with fidelity
 - Applicant submits at least one randomized control study (RCT) or quasi-experimental evaluation (QED) of the intervention the applicant will replicate
 - The evaluation found positive results for the intervention the applicant will replicate
 - The evaluation was conducted by an independent entity external to the organization whose program was studied
 - Applicant describes how the intervention studied and applicant's approach are the same
 - Applicant describes how they will replicate the intervention with fidelity to the program model
 - May be true but not required: Applicant has submitted a process evaluation demonstrating how it is currently replicating the intervention with fidelity to the program model
- **Preliminary – Option 3: Applicant has conducted at least one outcome study of its own intervention AND is proposing to replicate another evidence-based intervention with fidelity. All requirements outlined in Options 1 and 2 are met.**



Applicants proposing to replicate an evidence-based program with fidelity must describe how their program is the same as, or very similar to, the program they will replicate in the following areas:

- Characteristics of the beneficiary population
- Characteristics of the population delivering the intervention
- Dosage (frequency and duration) and design of the intervention
- Training for the AmeriCorps members and/or other individuals, such as volunteers, delivering the intervention
- The context in which the intervention is delivered
- Outcomes of the intervention

Applicants must also describe how they will assess whether they are implementing the intervention with fidelity to the intervention they are replicating.

Preliminary Evidence – Example 2 – Replication with fidelity

Applicant D will replicate the successful Money Matters financial literacy program. Money Matters utilizes trained volunteers to deliver a standardized financial literacy curriculum, paired with bi-weekly one-on-one coaching focused on setting one or two financial goals and taking small steps each month to meet the goal. A 2012 quasi-experimental study of Money Matters found that a year after completing the program, participants were significantly more likely than individuals in the comparison group to have a household budget, a checking account, and to have deposited money into a savings account within the past six months. Applicant D will replicate Money Matters with fidelity, providing the same training to AmeriCorps volunteers and using the same curriculum and coaching structure with program participants. We will collect output data from all sites to ensure that members complete all required training and that participants receive the intended dosage. A consultant from Money Matters will assist in training AmeriCorps members and will train site supervisors to conduct fidelity checks to ensure that the curriculum and coaching sessions are being implemented with fidelity.



What's good about this example:

- Describes the program that is being replicated and the results of the QED
- Describes how the Money Matters program will be replicated with fidelity (same population—volunteers—same training, standardized curriculum and coaching model, regular data collection and fidelity checks)

What could be improved:

- Dosage is not specified (may be in the applicant's logic model)
- Not clear if the Money Matters QED was conducted by an external evaluator
- Methodological quality & strength of findings will be determined through review of evaluation report, but the applicant should provide a short summary in the narrative as well

Evidence Tiers: Moderate

- **Moderate evidence**

- Applicant has conducted at least one quasi-experimental study (QED) or randomized control trial (RCT) of its own program
 - The studies are well-designed and well-implemented
 - The studies evaluate the same intervention described in the application
 - The studies demonstrate evidence of effectiveness (positive findings) on one or more key desired outcomes of interest depicted in the applicant's logic model
 - The studies were conducted by an independent entity external to the applicant organization
 - The ability to generalize the findings from the RCT or QED beyond the study context may be limited (e.g., single-site)

Moderate Evidence

In 2014, Applicant E's Ready to Read program conducted a randomized control trial at one of its fourteen sites. The study was conducted by an independent (external) evaluator. Students in the program outperformed students in the control group on reading comprehension. The effect size was moderate.

Evidence Tiers: Strong

- **Strong evidence**
 - Applicant has conducted at least one quasi-experimental study (QED) or randomized control trial (RCT) of its own program
 - The studies are well-designed and well-implemented
 - The studies evaluate the same intervention described in the application
 - The studies were conducted by an independent entity external to the applicant organization
 - The overall pattern of study findings is consistently positive
 - Findings from the studies may be generalized beyond the study context
 - At least one of the following is true:
 - The intervention has been tested nationally, regionally, or at the state level (e.g., multi-site) using a well-designed and well-implemented QED or RCT
 - The applicant has conducted multiple QEDs or RCTs in different locations or with different populations within a local geographic area

Strong Evidence

In 2014, Applicant E's Ready to Read program conducted a randomized control trial at all twenty-five sites statewide. The study was conducted by an independent (external) evaluator. By the end of the year, students in all sites in the Ready to Read program outperformed students in the comparison group on all literacy skills addressed by the program. The effect sizes were not only significant but substantial in magnitude. The Ready to Read program was effective regardless of gender, race/ethnicity, or dual language learner status and across multiple sites and site types.



Tips for Applicants

- Read all FAQs
- Present high quality evidence from the two strongest, most relevant studies
- Remember you may be considered for a lower evidence tier than the one you have self-assessed
- Describe the complete body of evidence that exists for your program
- Even if you submit studies, describe them in the narrative. The narrative and documents will be reviewed by different reviewers
- Respond to *all* elements within each selection criteria



Technical Assistance Webinars

Best Practices for Demonstrating Evidence

- Friday, April 8, 2:00 – 3:00 pm

California Performance Measurement Worksheets (PMWs)

- Wednesday, March 30, 2:00 – 3:00 pm
- Friday, April 8, 10:00 – 11:00 am

Q & A Sessions

- Wednesday, April 13, 10:00 – 11:00 am
- Friday, April 15, 2:00 – 3:00 pm



Questions?

- Email any questions to Funding@CaliforniaVolunteers.ca.gov
- FAQs will be posted weekly on CV's website: www.californiavolunteers.org/index.php/Grants/americanicorps/
- Additional resources can be found on CV's website

Thanks!

Notice of Intent
due **Friday, April 1, 2016**

Final Application
due **Tuesday, April 26, 2016**

