

NORC at the University of Chicago

# Overview of Evaluation Designs

# Learning Objectives



**By the end of this presentation, you will be able to:**

- Explain evaluation design
- Describe the differences between types of evaluation designs
- Identify the key elements of each type of evaluation design
- Understand the key considerations in selecting a design for conducting an evaluation of your AmeriCorps program

# Overview of the Presentation



- What is evidence and evaluation?
- What is evaluation design?
- AmeriCorps' evaluation continuum
- How to select an appropriate evaluation design for your program
- Key elements of each type of evaluation design
- Evaluation resources and tools

# What is Evidence?



- “The available body of facts or information indicating whether a belief or proposition is true or valid”

--OMB Circular No. A-11 Section 200 page 13

# Why is Evidence Important?



- Evidence generates credible, relevant, and actionable information about [AmeriCorps'] organizational effectiveness, operational performance, and the outcomes of national service programs
- Building evidence for what works, for whom, and under what circumstances, is a central part of ensuring the public's access to effective solutions.
- Evidence supports innovation, improvement, and learning

--AmeriCorps Strategic Learning and Evidence Building Plan, 2022-2026

# What is Evaluation?



- Evaluation is the use of research methods to assess a program's design, implementation, outcomes, or impacts.
- Evaluation looks at the results of your investment of time, expertise, resources, and energy, and compares those results with what you said you wanted to achieve in your program's logic model.

# Why are Evaluation and Evidence Building Important?



- Evaluation is a key driver of AmeriCorps program development, implementation, and continuous quality improvement
- Evaluation enables programs to systematically build evidence of effectiveness

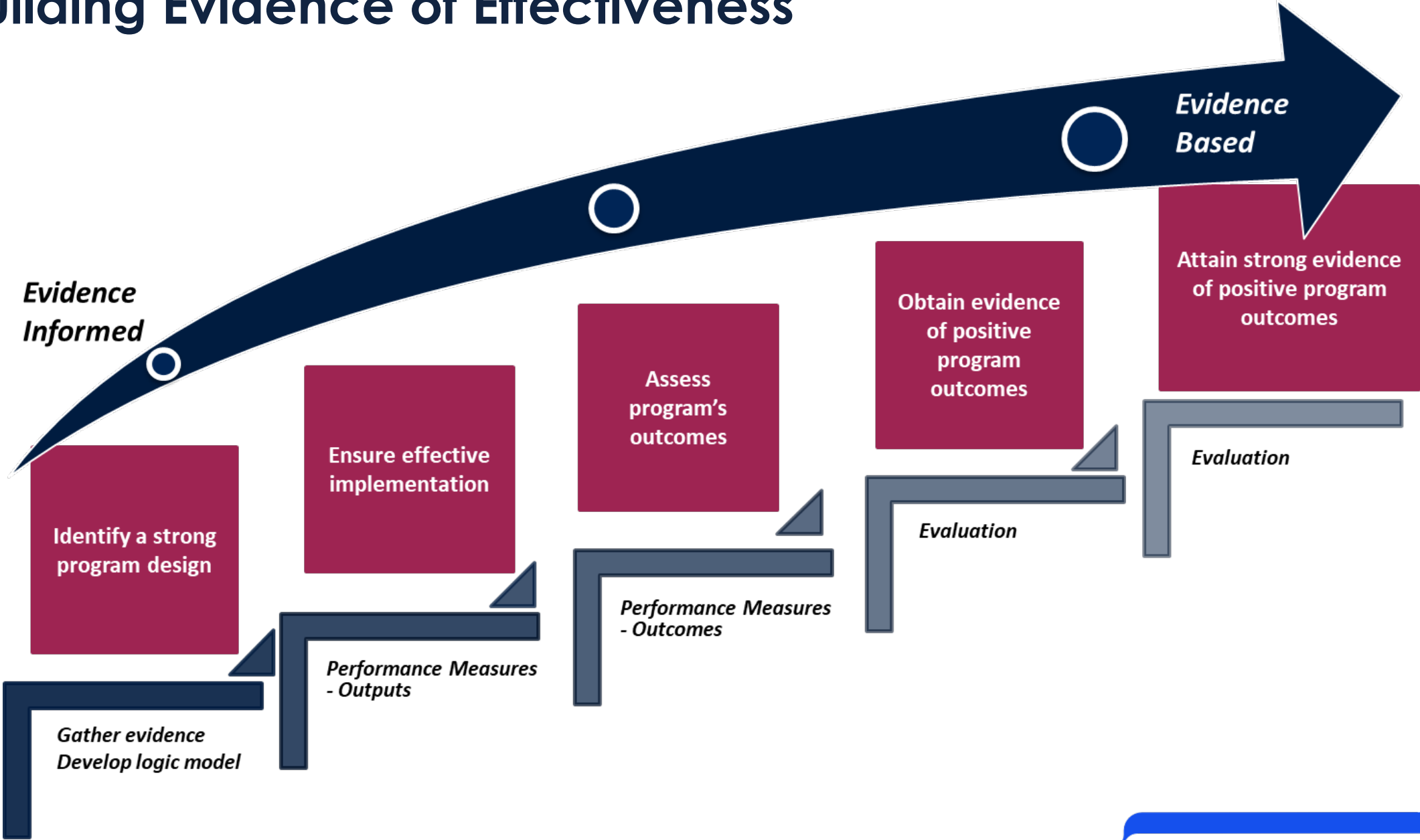
# Performance Measurement and Program Evaluation



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



# Building Evidence of Effectiveness





# What is Evaluation Design?

- Evaluation design is the structure that provides information to answer questions you have about your program. Evaluation design means thinking about:
  - Why conduct an evaluation
  - What to measure
  - Who to include in the evaluation (e.g., all beneficiaries or a sample)
  - When and how often data will be collected
  - What methods will be used to collect data
  - Whether comparison with another group is appropriate and feasible
- The evaluation design you choose depends on what kinds of questions your evaluation is meant to answer.



# Key Considerations in Selecting a Design

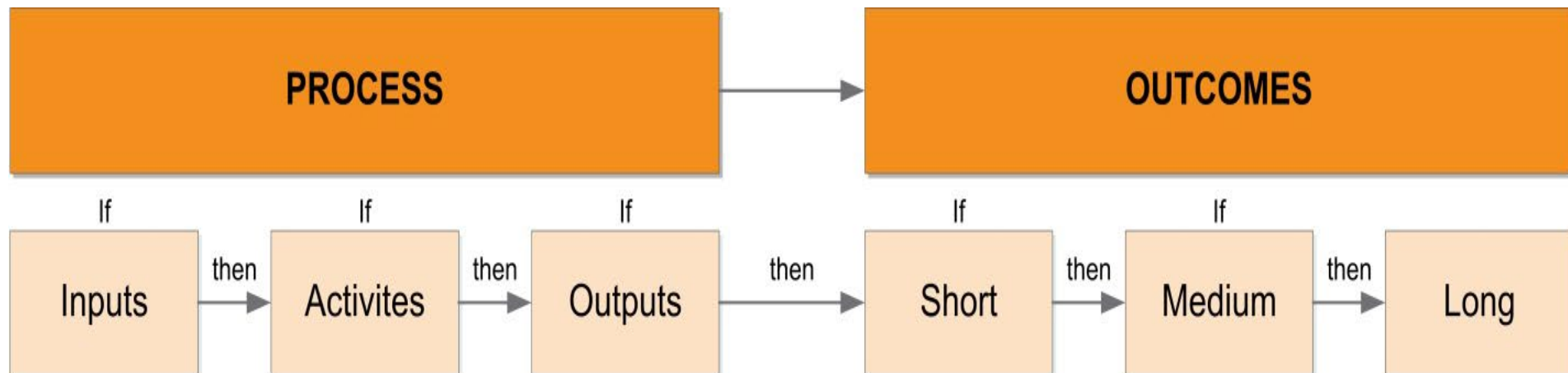
The appropriate design will largely depend upon:

- Your program model
- The primary purpose or goal of the evaluation
- The specific question(s) the evaluation will address
- Resources available for the evaluation
- Funder evaluation requirements

# Program Logic Model



- A program logic model is a detailed visual **representation of a program and its theory of change.**
- It communicates how a program works by depicting the intended relationships among program components.



# Example Logic Model for a Literacy Program



INPUTS	ACTIVITIES	OUTPUTS	Outcomes		
			Short-Term	Medium-Term	Long-Term
What we invest	What we do	Direct products from program activities	Changes in knowledge, skills, attitudes, opinions	Changes in behavior or action that result from participants' new knowledge	Meaningful changes, often in their condition or status in life
Funding Staff 100 AmeriCorps State and National members 75 non-AmeriCorps volunteers Research	One-on-one tutoring to students below benchmark	Number of students receiving tutoring assistance	Increase in number of students scoring at or above benchmark on literacy assessments Improved student self-efficacy	Increase in number of students reading on grade-level	Students maintain grade-level proficiency in reading

For an overview of logic models, AmeriCorps grantees can refer to the module, “How to Develop a Program Logic Model” located on the Evaluation Resources page.

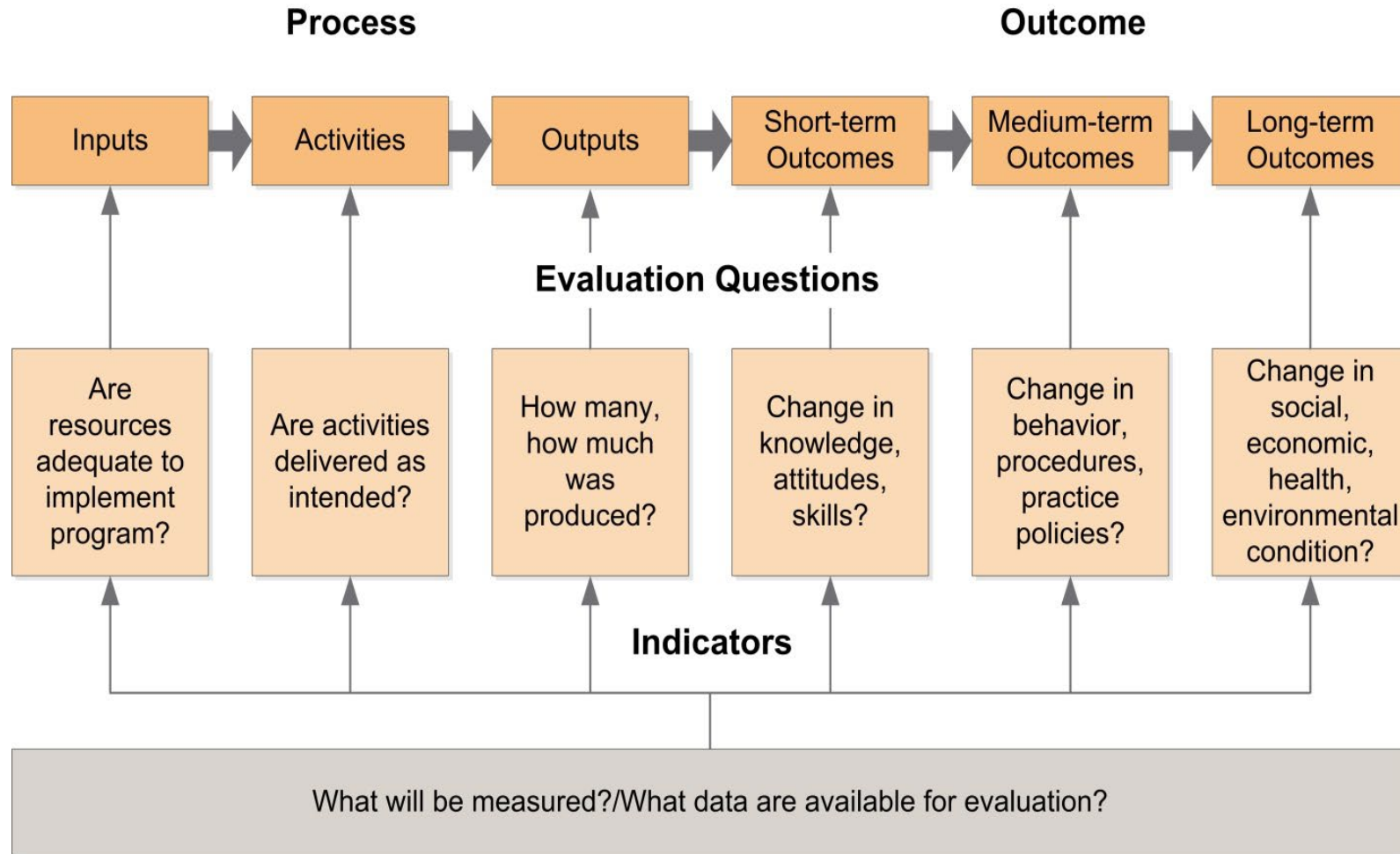
# Define Purpose and Scope



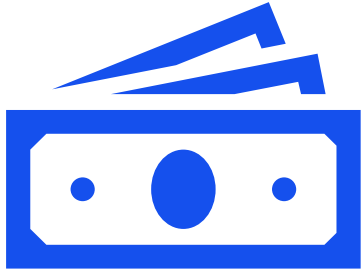
Each evaluation should have a primary purpose around which it can be designed and planned.

- Why is the evaluation being done?
- What do you want to learn?
- How will the results be used? By whom?

# Selecting Research Questions



# Resource Considerations

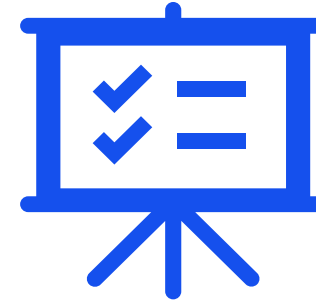


**Consider what resources are available to carry out the evaluation:**

Staff time

Funding

Outside evaluation expertise



**It is not necessary to evaluate your entire program.**

Evaluation can be narrow or broad depending on questions to be answered

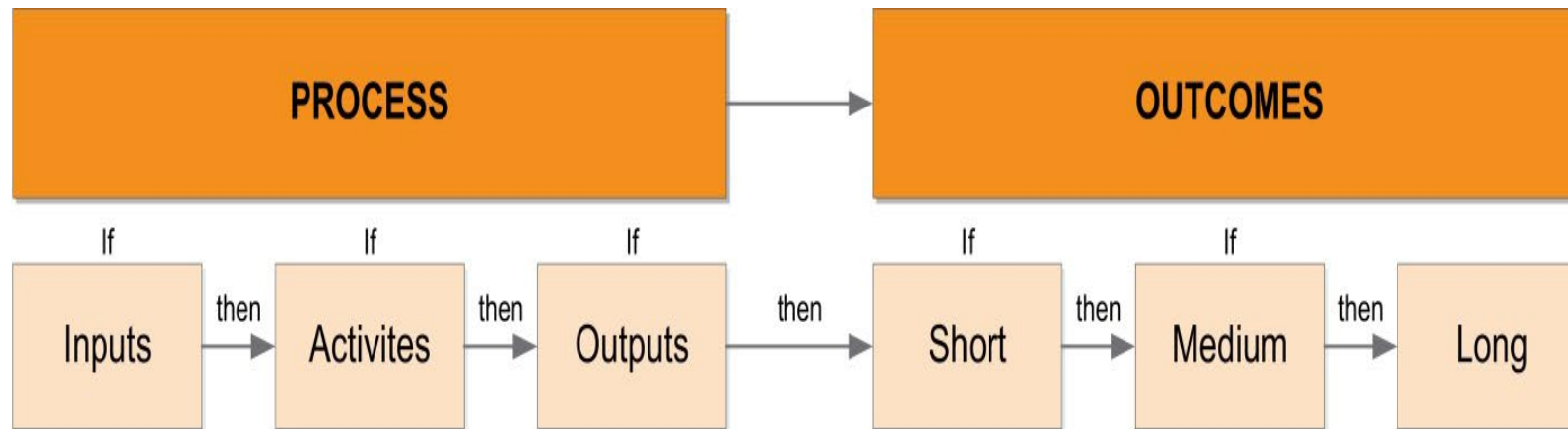
Evaluation is not a one-time activity but a series of activities over time that align with the life cycle of your program



# Basic Types of Evaluation Designs



- The two “sides” of a program’s logic model align with the two types of evaluation designs: Process and Outcome.



# Basic Types of Evaluation Designs



Description of general categories of evaluation designs:

Process Evaluation	Outcome Evaluation	Impact Evaluation
<ul style="list-style-type: none"><li>• Examines the extent to which a program is operating as intended by assessing ongoing program operations and determining whether the target population is being served</li><li>• Results may be used to determine what changes and/or improvements should be made to the program's operations</li></ul>	<ul style="list-style-type: none"><li>• Measures changes in knowledge, attitude(s), behavior(s) and/or condition(s) among program beneficiaries or other stakeholder groups</li><li>• Results may demonstrate what the program has achieved</li></ul>	<ul style="list-style-type: none"><li>• Measures changes in program beneficiaries or stakeholders before and after an intervention relative to a reasonably similar comparison/control group</li><li>• Results are an estimate of the program's impact on beneficiaries or other stakeholder groups</li></ul>

# AmeriCorps Approved Evaluation Designs



Type of Design	Category	Details needed on evaluation design
Process	Process	<ul style="list-style-type: none"> <li>Description of the methods that will be used (i.e., qualitative only, quantitative only, or mixed methods)</li> </ul>
Non-experimental Design	Outcome	<ul style="list-style-type: none"> <li>Description of whether pre- AND post-test measurements OR post-only measurements will be used</li> </ul>
Quasi-experimental Design (QED)	Impact	<ul style="list-style-type: none"> <li>Description of the approach for identifying a reasonably similar comparison group (e.g., propensity score matching, difference in difference analysis)</li> <li>List of variables (covariates) to be used to statistically equate treatment and comparison groups at baseline</li> </ul>
Experimental Design/Randomized Controlled Trial (RCT)	Impact	<ul style="list-style-type: none"> <li>Description of the random assignment procedures that will be used to form treatment and control groups</li> <li>Eligibility criteria for inclusion in the study</li> </ul>

# Process Evaluation



- Goals:
  - Documents what the program is doing
  - Documents to what extent and how consistently the program has been implemented as intended
  - Informs changes or improvements in the program's operations
- Common features:
  - Does not require a comparison group
  - Includes qualitative and quantitative data collection
  - Does not require advanced statistical methods



# Examples of Process Evaluation Questions

- Is the program being implemented as designed or planned?
  - Is the program being implemented the same way at each site?
  - Is the program reaching the intended target population with the appropriate services at the planned rate and "dosage"?
- Are there any components of the program that are not working well? Why or why not?
- Are program beneficiaries generally satisfied with the program? Why or why not?
- Are the resources adequate for the successful implementation of the program?

# Examples of Methods and Data Collection Tools for Process Evaluation



## Data Sources:

- Program and school level administrative data
- Site visits to the schools to examine the fidelity of program implementation
  - Observations of literacy intervention with individual students
  - Interviews with school staff and administration
  - Focus groups with teachers and students

## Analysis:

- Thematic identification
- Confirmation of findings across sources

# Group Exercise #1: Designing a Process Evaluation for a Literacy Program



## Research question:

- Is the literacy program being implemented consistent with the program's logic model and theory of change?

## Design considerations:

- What to measure
- Who to include in the evaluation
- When and how often data will be collected
- What methods will be used to collect data

# Group Exercise #1: Designing a Process Evaluation for a Literacy Program



Crosswalk for Process Evaluation of a Literacy Program				
Research question	Indicators	From whom / data sources?	When collected and by whom?	How will you analyze the data?
Is the literacy program being implemented as designed?				



# Example Crosswalk for a Process Evaluation of a Literacy Program



Crosswalk for Process Evaluation of a Literacy Program				
Research question	Indicators	From whom / data sources?	When collected and by whom?	How will you analyze the data?
Is the literacy program being implemented as designed?	<ul style="list-style-type: none"> <li>a) Member use of program curriculum during tutoring</li> <li>b) Duration of tutoring</li> <li>c) Student participation rates</li> </ul>	<ul style="list-style-type: none"> <li>a, b, c) Members</li> <li>a, b) Evaluator observes members tutoring students</li> </ul>	<ul style="list-style-type: none"> <li>a, b, c) Members report details of sessions in tutoring logs quarterly</li> <li>a) Quarterly observations by the evaluator(s) using structured observation protocols</li> </ul>	<ul style="list-style-type: none"> <li>a, b, c) Generate frequencies on use of curriculum; average duration of sessions; and average rate of student attendance</li> <li>c) Generate frequencies and averages on quantitative data (e.g., ratings scales, frequency scales) and thematically code and analyze open-ended comments/notes</li> </ul>

# Optional Exercise #1: Designing a Process Evaluation for a Literacy Program



Crosswalk for Process Evaluation of a Literacy Program				
Research question	Indicators	From whom / data sources?	When collected and by whom?	How will you analyze the data?
Are program beneficiaries generally satisfied with the program? Why or why not?				

# Example Crosswalk for a Process Evaluation of a Literacy Program



Crosswalk for Process Evaluation of a Literacy Program				
Research question	Indicators	From whom / data sources?	When collected and by whom?	How will you analyze the data?
<p>Are program beneficiaries generally satisfied with the program? Why or why not?</p>	<p>a) Satisfaction level b) Attendance</p>	<p>a) Parents, teachers, administrators b) Students</p>	<p>a) Parent survey sent home at end of tutoring; teacher focus groups and administrator interviews completed by evaluator b) Members record attendance in tutoring logs quarterly</p>	<p>a) Generate averages on quantitative survey data (e.g., ratings scales, frequency scales) and thematically code and analyze responses from focus groups and interviews b) Generate average rate of student attendance</p>



# Outcome Evaluation

## Goals:

- Identifies the results or effects of a program
- Measures program beneficiaries' changes in knowledge, attitude(s), and/or behavior(s) before and after program participation

## Common Features:

- Typically requires quantitative data
- May or may not use advanced statistical methods
- May or may not include a comparison group (only when collecting data at one time point after the intervention)

# Outcome Evaluation Questions



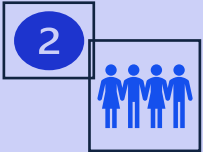
- Did the program beneficiaries experience a change in knowledge, attitude, behavior, or condition after participating in the program?
- Did all types of program beneficiaries benefit from the program or only specific subgroups?

# Outcome Evaluation Designs



## Single group pre-post design

Provides a comparison of program beneficiaries before and after they receive program services



## Two group post design

Examines outcomes for program beneficiaries after they receive program services and for a similar comparison group



## Retrospective study designs

Ask previous program beneficiaries to provide opinion on the effects of the program services they received



## Member surveys

Survey members on their program experiences and opinions of the results of their service

# Impact Evaluation



## Goals:

- Measures program impacts on program beneficiaries' knowledge, attitude(s), and/or behavior(s) relative to a comparison group

## Common Features:

- Requires quantitative data and advanced statistical methods
- Includes a comparison group

## Evaluation Question:

- Are there differences in outcomes for program beneficiaries before and after program participation compared to a similar group not in the program?

# Impact Evaluation Designs



- Quasi-experimental design (QED)
  - Form comparison group from a similar population of program participants (e.g., similar participants from another program, extra applicants, etc.)
- Experimental design (Randomized Controlled Trial- RCT)
  - Randomly assign new eligible applicants to either receive intervention/program or alternative/delayed services (control group)





# Quasi-Experimental vs. Experimental Designs

Quasi-Experimental	Experimental
<ul style="list-style-type: none"><li>• Can be challenging to identify a similar comparison group</li><li>• Because program and comparison groups are different, results are considered less rigorous</li><li>• Often less labor intensive and expensive than experimental design</li></ul>	<ul style="list-style-type: none"><li>• Most rigorous design option, so results tend to be more highly regarded</li><li>• Often requires increased program recruitment</li><li>• Applicant acceptance is randomly determined</li><li>• Can be more labor intensive and expensive</li></ul>

# What is a Comparison or Control Group?



- A group of individuals not participating in the program or receiving the intervention
- Necessary to determine if the program, rather than some other factor, is causing observed changes
- “Comparison group” is associated with a quasi-experimental design and “control group” is associated with an experimental design

# Other Types of Quasi-experimental Designs



- AmeriCorps approves some other types of QED evaluation designs for large, recompetete grantees without the use of matching/propensity scoring if a reasonable comparison group is identified
- Examples: Interrupted Time Series Design, Regression Discontinuity Design, etc.

# Group Exercise #2: Designing an Impact Evaluation of a Literacy Program



## Research question:

What impact does the literacy intervention program have on student reading levels relative to a comparison group of similar students?

## Design considerations:

- What to measure
- Who to include in the evaluation
- When and how often data will be collected
- What methods will be used to collect data

# Group Exercise #2: Designing an Impact Evaluation for a Literacy Program



Crosswalk for Impact Evaluation of a Literacy Program				
Research question	Indicators	From whom / data sources?	When collected and by whom?	How will you analyze the data?
What impact does the literacy intervention program have on student reading levels relative to a comparison group of students?				

# Example Crosswalk for an Impact Evaluation of a Literacy Program



Crosswalk for Impact Evaluation of a Literacy Program				
Research question	Indicators	From whom / data sources?	When collected and by whom?	How will you analyze the data?
What impact does the literacy intervention program have on student reading levels relative to a comparison group of students?	Student reading achievement is measured with literacy assessment tests.	Students participating in the program serve as the intervention group. Students enrolled at a similar school with no program serve as the comparison group.	The evaluator administers the assessments at two time points: <ul style="list-style-type: none"> <li>- At the beginning of the school semester</li> <li>- At the end of the school semester</li> </ul>	Calculate an Analysis of Variance (ANOVA) or regression analysis to determine if there are any statistical differences in the average difference in outcomes between the intervention group and the comparison group before and after treatment

# Optional Exercise #2: Designing an Impact Evaluation for a Literacy Program



Crosswalk for Impact Evaluation of a Literacy Program				
Research question	Indicators	From whom / data sources?	When collected and by whom?	How will you analyze the data?
What impact does the literacy intervention program have on student self-efficacy relative to a comparison group of students?				

# Example Crosswalk for an Impact Evaluation of a Literacy Program



Crosswalk for Impact Evaluation of a Literacy Program				
Research question	Indicators	From whom / data sources?	When collected and by whom?	How will you analyze the data?
What impact does the literacy intervention program have on student self-efficacy relative to a comparison group of students?	Student self-efficacy is measured with existing tools, such as the self-efficacy questionnaire for children (SEQ-C).	Students participating in the program serve as the intervention group. Students enrolled at a similar school with no program serve as the comparison group.	The evaluator administers the assessments at two time points: <ul style="list-style-type: none"> <li>- At the beginning of the school semester</li> <li>- At the end of the school semester</li> </ul>	Calculate the difference in average outcome in the intervention group minus the difference in average outcome in the comparison group before and after treatment (difference in differences method)



# Small and Large Requirements for an Evaluation Plan



	Small Grantees (<\$500,000)	Large Grantees (>\$500,000)
When are grantees required to conduct an evaluation?	During 2 <sup>nd</sup> competitive grant cycle, and all following competitive cycles.	During 2 <sup>nd</sup> competitive grant cycle, and all following competitive cycles.
When must evaluation plans be approved?	By the end of the first year of the second grant cycle.	By the end of the first year of the second grant cycle.
What evaluation design is required?	Any design (e.g., process, non-experimental outcomes).	Impact evaluation (quasi-experimental or experimental)
What type of evaluator?	Internal or external evaluator.	External evaluator.

# Alternative Evaluation Approach (AEA)



AEA	Grantee	Justification
Funding threshold	Large	- Grantees who receive an average of less than \$1 million per year can request to be exempt from the large grantee requirements and conduct an internal non-impact evaluation.
Previous impact evaluation	Large	- Previously conducted an impact evaluation with demonstrated evidence of effectiveness (i.e., Strong or Moderate evidence).
AmeriCorps National Evaluation	Large or Small	- Grantees participating in an AmeriCorps' Office of Research and Evaluation national evaluation (i.e., bundled evaluation or Return on Investment) that will not be completed during current grant cycle - Large grantees can also request this if the national evaluation's design does not fulfill the requirements for a large grantee.

# Alternative Evaluation Approach (AEA) – continued



AEA	Grantee	Justification
Structure of program or grantee organization	Large	<ul style="list-style-type: none"><li>- Insurmountable challenges forming a comparison group.</li><li>- Significant changes to program design.</li></ul>
Replication	Large	<ul style="list-style-type: none"><li>- Implementing an evidence-based intervention with fidelity in a new setting.</li><li>- A grantee's application must be assessed at the Strong or Moderate evidence level.</li></ul>
Timing	Large or Small	<ul style="list-style-type: none"><li>- Evaluation will not be completed by end of current grant cycle.</li><li>- AEA approval required only if an interim evaluation report will not meet evaluation requirements.</li></ul>



# Internal Resources

- Resources for State and National Direct Grantees  
[https://americorps.gov/grantees-sponsors/directs-territories-tribes?field\\_document\\_type\\_tax\\_target\\_id=19756#resources](https://americorps.gov/grantees-sponsors/directs-territories-tribes?field_document_type_tax_target_id=19756#resources)
- This page includes information about the following:
  - AEA Guidance and Request Form
  - Evaluation FAQs
  - Evaluation Plan Template
  - Evaluation Requirements
- ASN – Evaluation Resources  
<https://www.americorps.gov/grantees-sponsors/evaluation-resources>
- AmeriCorps Evaluation TA Portal
- <https://americorpsevaluationta.norc.org>



## External Resources

- The American Evaluation Association: <http://www.eval.org>
- The Evaluation Center: <http://www.wmich.edu/evalctr/>
- Innovation Network's Point K Learning Center: <http://www.innonet.org>

# Questions and Answers



# Thank you!

Carrie E. Markovitz, Ph.D.

NORC at the University of Chicago

[markovitz-carrie@norc.org](mailto:markovitz-carrie@norc.org)

To contact the Office of Research and  
Evaluation: [evaluation@cns.gov](mailto:evaluation@cns.gov)