






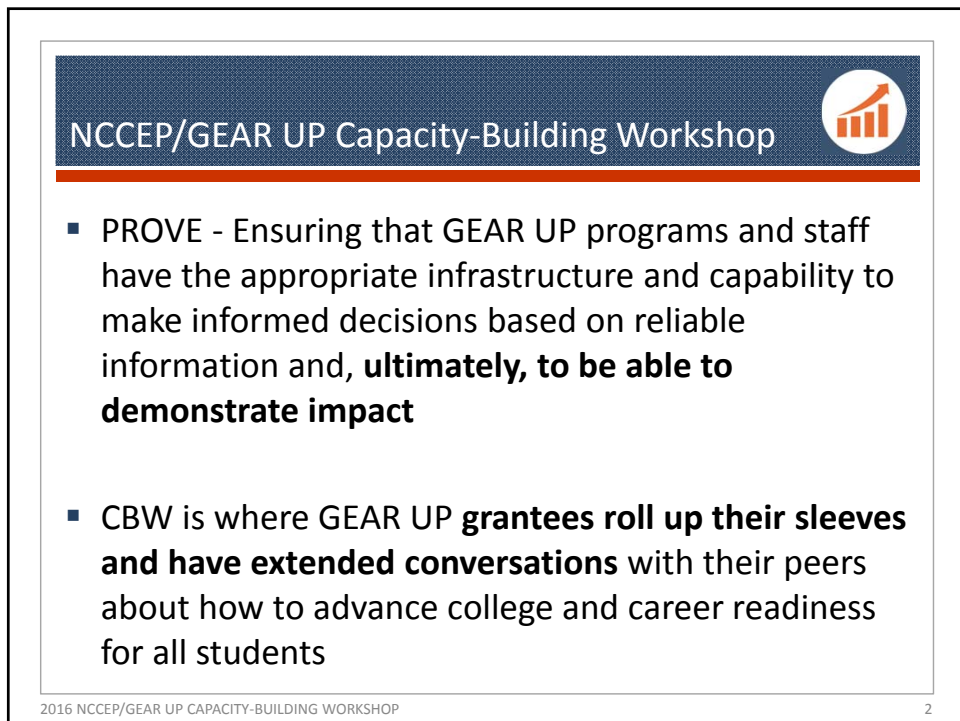
New Orleans, Louisiana  
Hilton New Orleans Riverside  
Jan. 31 - Feb. 3, 2016




NCCEP/GEAR UP EXCEL • PROVE • MOBILIZE  
CAPACITY-BUILDING WORKSHOP

PROVE WORKSHOP:  
*P5: Applying Rigorous Research*

Alina Martinez, Abt Associates Inc.



NCCEP/GEAR UP Capacity-Building Workshop 

- PROVE - Ensuring that GEAR UP programs and staff have the appropriate infrastructure and capability to make informed decisions based on reliable information and, **ultimately, to be able to demonstrate impact**
- CBW is where GEAR UP **grantees roll up their sleeves and have extended conversations** with their peers about how to advance college and career readiness for all students

2016 NCCEP/GEAR UP CAPACITY-BUILDING WORKSHOP 2

## PROVE – Building Rigorous Evidence



### Today:

- Need for rigorous research and evaluation, in order to build a strong evidence base within GEAR UP
- Ideas for moving in this direction

## Building Rigorous Evidence



### Objectives:

- Gain an understanding of trends in research and evaluation as we move towards increased evidence-based outcomes in college access and success.
- Learn about techniques that can be applied to your current evaluation or in a new grant application to increase evidence-based outcomes.
- Learn about the application of techniques for increasing rigorous research and evaluation in GEAR UP programs.

## Workshop Overview



- What came before (evaluation in GEAR UP)
- Federal expectations regarding evidence
- ED's evidence levels
- What Works Clearinghouse Standards and their relationship to ED's evidence level
- Key components of rigorous evaluation

## Continuing to Build Rigorous Evidence



### Raising the Bar on GEAR UP Evaluation:

#### Implementing Experimental Design and Random Assignment Studies

A participative panel discussion hosted by  
Keren Brooks CoBro Consulting  
Chrisy Tillery NCEP  
Kurt R. Burkum ACT  
Changhua Wang Education Northwest  
Janet Usinger University of Nevada, Reno

2014 NCEP/GEAR UP Annual Conference

### BUILDING A BODY OF EVIDENCE ON THE EFFECTS OF GEAR UP

Marsha Silverberg  
Institute of Education Sciences  
National Center for Education Evaluation

## Evaluation in GEAR UP – Strong Plan (2014 RFP)



- A strong **evaluation plan** should:
  - Shape the development of the project from the beginning to the end of the grant period;
  - Include benchmarks to monitor progress toward meeting specific project objectives;
  - Include a summative evaluation approach for assessing the likely contribution of the project to improving student outcomes.

## GEAR UP Evaluation Plans (2014 RFP)



- Plan should clearly indicate:
  1. What types of data will be collected;
  2. When various types of data will be collected;
  3. For which students data will be collected – participants only, participants and non-participants, which cohort (i.e., in what year did the students start);
  4. What data collection methods will be used;
  5. How the data will be analyzed; and
  6. When reports and outcome data will be available.

## GEAR UP Evaluation Plans (2014 RFP)



- Plans should maximize how much confidence the Project or the Program can have in the study results
- Strongest evaluation designs are experimental/randomized control studies
- Other evaluation designs may be easier to implement but do not provide such rigorous results.

## Evaluation Independence



- The evaluation should be conducted by an evaluator that is independent of the entities developing or implementing the intervention.
- Independence is important in various aspects of the evaluation, including:
  - Collection of outcome data
  - Analysis of impacts
  - Reporting of the estimated effects of the intervention

## Activity – Your Project’s Evaluation Plans



- Discuss at your table
  - What do/did you want to learn from your evaluation?
  - What are/were the key components of the evaluation design?
- Share with full group
  - Areas of investigation
  - Key components
    - Similarities
    - Differences

## Activity – Reporting Out



- Areas of investigation
- Key components of evaluations
  - Similarities
  - Differences

## Some Issues Important to Rigorous Evaluation



### We will tackle a few

- Construction of comparison group
- Attrition
- Baseline equivalence
- Appropriate outcomes
- Confounds
- And point to resources for others
  - Logic models
  - Implementation fidelity
  - Power analyses

## Why Does Rigor Matter?



- Increasing federal emphasis on rigorous evaluation
- Providing evidence of effectiveness
- Ensuring that federal dollars are invested wisely
- Informing continuous improvement of programs
  - Change what is not working
  - Do more of what is working

## Federal Context

Presidential Administrations	Federal Guidance
President Clinton (1993 – 2001)	Government Performance and Results Act of 1993 (GPRA)
President Bush (2001 – 2009)	Program Assessment Rating Tool (PART)
President Obama (2009 – 2017)	<ul style="list-style-type: none"> <li>• GPRA Modernization Act of 2010</li> <li>• Office of Management and Budget Memoranda</li> </ul>


2016 NCCEP/GEAR UP CAPACITY-BUILDING WORKSHOP 15

## Convergence of Efforts to Strengthen Federal Research and Evaluation

- Office of Management and Budget (OMB)
- Congress
- Focus on evidence across Agencies
- ED efforts
  - Common guidelines for education research and development (Aug 2013)  
<http://ies.ed.gov/pdf/CommonGuidelines.pdf>
  - What Works Clearinghouse™
  - Evaluation technical assistance to grants

2016 NCCEP/GEAR UP CAPACITY-BUILDING WORKSHOP 16






## OMB Memoranda

- M-10 -01 Increased Emphasis on Program Evaluation
- M-12-14 Use of Evidence and Evaluation in the 2014 Budget
- M-13-17 Next Steps in the Evidence and Innovation Agenda
- M-14-07 Fiscal Year 2016 Budget Guidance, Evidence and Evaluation

2016 NCEP/GEAR UP CAPACITY-BUILDING WORKSHOP
17



## Focus on Evidence across Agencies

Department or Agency	Example Initiative
Education	Investing in Innovation (i3)
Health and Human Services	Maternal, Infant, and Early Childhood Home Visiting Program
Labor/Education	Trade Adjustment Assistance Community College and Career Training (TAACCT)
Labor/Education/Health and Human Services	Workforce Innovation Fund
Corporation for National and Community Service	Social Innovation Fund

2016 NCEP/GEAR UP CAPACITY-BUILDING WORKSHOP
18

## Education Department General Administrative Regulations (EDGAR)



- Identifies standards for evidence for ED
- Specifies requirements for all ED federal grants


Criteria used to evaluate GEAR UP applications are drawn from EDGAR 34 CFR part 75 sections 75.209(a) and 75.210.

## Relevance of EDGAR to RFPs



- 7 Selection Criteria Stemming from EDGAR
  1. Need for the project (15 points)
  2. Quality of project design (15 points)
  3. Quality of project services (15 points)
  4. Quality of project personnel (10 points)
  5. Quality of the management plan (10 points)
  6. Quality of the project evaluation (20 points)
  7. Adequacy of project resources (15 points)
- In addition, there were competitive preference priorities (3 for States, 2 for Partnerships)

## Simplified Overview of EDGAR Evidence Categories



Strong theory


Evidence of promise

Moderate evidence of effectiveness

Strong evidence of effectiveness

2016 NCCEP/GEAR UP CAPACITY-BUILDING WORKSHOP
21



## ED Efforts – Useful Resources



### Common Guidelines

[Common Guidelines for Education Research and Development](#)

A Report from the Institute of Education Sciences, U.S. Department of Education and the National Science Foundation  
 August 2013

### What Works Clearinghouse™

ies INSTITUTE OF EDUCATION SCIENCES
WHAT WORKS CLEARINGHOUSE™

**WWC: Who, What, and Why**

**Who we are:**


- ▶ An initiative of the Institute of Education Sciences (IES) at the U.S. Department of Education

**What we do:**

- ▶ We review the research in education to answer the question "What works in education?"

**Why we do it:**

- ▶ To provide educators with the information to make evidence-based decisions



2016 NCCEP/GEAR UP CAPACITY-BUILDING WORKSHOP
22

## WWC Standards




- Focus on causal validity of the study design and analyses
- Identify standards for rigorous evidence
- Developed by panels of national experts
- Result in rating after study review of
  - Meets WWC Standards without Reservations
  - Meets WWC Standards with Reservations
  - Does Not Meet Standards

## Definitions of Summative Evaluation Design Approaches (RFP)




- Experimental study
- Quasi-experimental study
- Carefully matched comparison group design

**Definitions of Summative Evaluation Design Approaches(RFP)** 

- **Experimental study**

A study that employs random assignment of, for example, students, teachers, classrooms, schools, or districts to participate in a project being evaluated (treatment group) or not to participate in the project (control group). The effect of the project is the average difference in outcomes between the treatment and control groups.

2016 NCCEP/GEAR UP CAPACITY-BUILDING WORKSHOP 25

**Definitions of Summative Evaluation Design Approaches(RFP)** 

- **Experimental study**
- **Quasi-experimental study**

An evaluation design that attempts to approximate an experimental design and can support causal conclusions (i.e., minimizes threats to internal validity, such as selection bias, or allows them to be modeled). Well-designed quasi-experimental studies include carefully matched comparison group designs, interrupted time series designs, or regression discontinuity designs.

2016 NCCEP/GEAR UP CAPACITY-BUILDING WORKSHOP 26

## Definitions of Summative Evaluation Design Approaches(RFP)



- Experimental study
- Quasi-experimental study
- Carefully matched comparison group design

A type of quasi-experimental study that attempts to approximate an experimental study. More specifically, it is a design in which project participants are matched with non-participants based on key characteristics that are thought to be related to the outcome.

## Activity – Constructing a Comparison Group



- How would you construct a comparison group to evaluate:
  - Your entire GEAR UP project?
  - A component of your project?

## Activity – Reporting Out



- How many RCTs were proposed for:
  - Entire GEAR UP project
  - Components of projects
- What were some approaches to constructing a non RCT comparison group?

## Rigor of Design for Causal Inference



- Designs considered rigorous enough to provide strong evidence of effectiveness (All of these designs have a comparison group)
  - Randomized control trial (RCT)
  - Quasi-experimental design (QED)
  - Regression discontinuity design (not a topic for today –resource sheet)
  - Comparative interrupted time series (not a topic for today–resource sheet)
- Designs not rigorous enough to provide evidence of effectiveness (These designs do not have a comparison group)
  - Interrupted time series
  - Pre-post with no comparison group
  - Correlational studies
  - Post-test only

## RCT vs QED

RCTs	QEDs
<b>Random mechanism assigns volunteers to participant or non-participant control group</b>	Non-random mechanism identifies participants and matches them to a non-participant comparison group on pre-program characteristics
<b>Groups are similar on both observed and unobserved characteristics</b>	Groups are matched as closely as possible on any observable/measurable characteristic that could affect outcomes  Groups may differ on unobserved characteristics
<b>Differences in outcome can be attributed to the program/strategy/practice</b>	Differences in outcomes may be due to the intervention and other factors unobserved

2016 NCEP/GEAR UP CAPACITY-BUILDING WORKSHOP 31

## Important Considerations/Components


- **Logic model**
- **Fidelity of implementation**
- **Comparison group**
  - Distinct groups
  - Baseline equivalence
- **Attrition**
- **No confounding factors**
- **Appropriate outcome**
- **Sufficient power (consider level and units)**

2016 NCEP/GEAR UP CAPACITY-BUILDING WORKSHOP 32

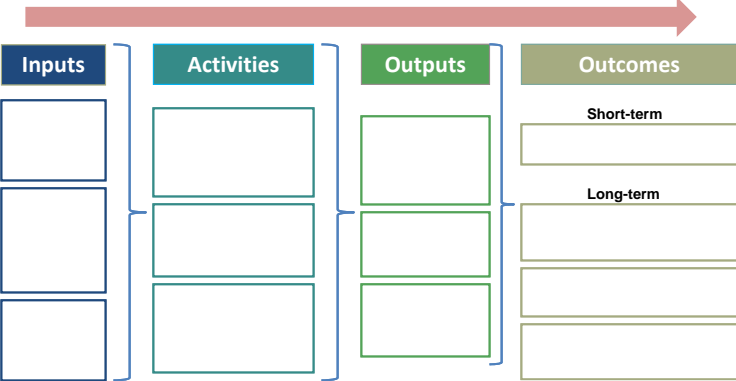


### Logic Model

You need to know what “it” is




**GOAL:**



2016 NCEP/GEAR UP CAPACITY-BUILDING WORKSHOP 33

### Logic Model

Understanding the “it”



- Implementing multiple strategies (i.e. interventions) concurrently poses challenges to measuring the impact of any of them.
- Need to clearly identify what is being implemented, for whom, how it’s different from “business as usual.”
- Use a logic model.

National Evaluation of i3 TA resources  
[email NEi3\_TA@abtassoc.com]

2016 NCEP/GEAR UP CAPACITY-BUILDING WORKSHOP 34

## Implementation Fidelity

Evidence that 'it' was implemented.



- Measuring implementation fidelity is important...
- ...but impossible to do without a clear understanding of how the intervention is intended to work!
- It is also important to understand whether participants in the comparison condition are having experiences that look at all similar to the treatment condition.

National Evaluation of i3 TA resources  
[[email NEi3\\_TA@abtassoc.com](mailto:NEi3_TA@abtassoc.com)]

## Comparison Group

Distinct Groups – (Two Groups, At Least)



- Compare treatment group to comparison group that does not receive the intervention
- Similar on everything except the intervention
- **Not** a comparison of treatment to itself (e.g. same students in previous grade)
- **Not** a previous or later cohort (e.g. freshmen from the previous year)

## Baseline Equivalence



- Treatment and comparison groups are similar on observed pre-intervention measures.
- Established on (1) a pre-intervention measure of the outcome or (2) one or more variables that are highly correlated with this outcome.
- Difference in means is  $< 0.25$  standard deviations.
  - Adjust statistically for differences  $> 0.05$  standard deviations.
  - No adjustment required for differences  $\leq 0.05$  standard deviations.

## Understanding Attrition



- Attrition occurs when the value of the outcome variable is missing for a unit (e.g., student) that was randomly assigned.
- The WWC has standards for the maximum acceptable attrition rate and the maximum acceptable difference in attrition rates between the treatment and control groups.

## Confounds

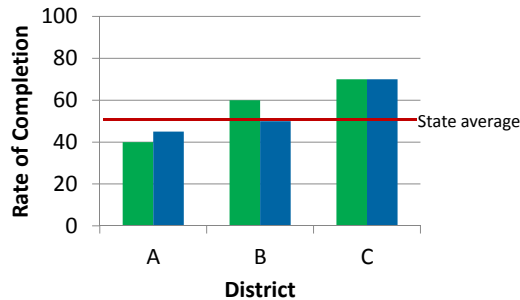


- **Confounds** are design features that may yield differences in outcomes between the two groups that are *not* attributable to the intervention.
- **Examples:**
  - An “N=1 problem”, in which there is only one “unit”—teacher, student, school, or district—per condition.
  - Systematic differences in sample characteristics that are associated with outcomes (e.g., teacher education).
  - Bundling of the intervention with another intervention.

## Why a comparison group matters




### FAFSA completion rate




In 2 of 3 districts, GEAR UP schools above the state's rate of 50%

In only 1 district are GEAR UP schools above that of comparable non-GEAR UP school

**Appropriate Outcomes** 

- Follow from theory of intervention
- Used to determine impacts
- Include
  - Face validity – appears to measure intended outcome
  - Reliability – consistency
  - Not overly aligned with intervention
  - Collected in same manner for treatment and comparison

2016 NCEP/GEAR UP CAPACITY-BUILDING WORKSHOP 41

**Outcomes:** 

**1. Face validity**

- An outcome has face validity if it appears to measure the domain into which it has been classified.

2016 NCEP/GEAR UP CAPACITY-BUILDING WORKSHOP 42

## Outcomes: 2. Reliability



- Minimum WWC standards for reliability:
  - Test-retest reliability  $\geq 0.40$
  - Internal consistency  $\geq 0.50$
  - Inter-rater reliability  $\geq 0.50$
- Standardized tests (e.g., nationally normed tests and state assessments) are assumed to have adequate reliability.

## Outcomes: 3. Alignment with the Intervention



- There should not be **over-alignment**.
- Over-alignment occurs when outcome measures are closely tailored to the intervention.
- Example: An outcome measure based on an assessment that relied on materials used in the intervention condition but not in the comparison condition (e.g., specific reading passages) likely would be judged to be overaligned.

## Outcomes:

### 4. Collected in the Same Manner



- Should be **collected in the same manner** for the treatment and comparison groups.
- A measure would be collected in a *different* manner, for example, if:
  - Different modes, timing, or personnel were used to collect outcome data for the two groups.
  - Outcome measures were constructed differently for the two groups.

## Power Analysis



- Find the needed sample size
  - For a given power and effect size
- Find the power
  - For a given sample size and effect size
- Find the minimum detectable effect size (MDE)
  - For given sample size and power

National Evaluation of i3 TA resources  
 [email NEi3\_TA@abtassoc.com]

## Discussion



- What are some common issues encountered in RCT and QEDs?

## Common Challenges



- RCT
  - Perception of denied services/unfairness
  - Recruiting participants
  - Information needed to implement random assignment (*what do you need and how will you get it?*)
  - Attrition
- QED
  - Constructing a comparison group
  - Matching
  - Establishing equivalence at baseline
  - Extant data on the comparison group
- Both
  - Collecting outcome data in the same way across T and C groups
  - Accounting for clustering



## Discussion – Moving toward More Rigor

- How do we collectively move to more rigorous accountability?
- What steps can you take to move your GEAR UP project toward more rigorous evaluation?

2016 NCEP/GEAR UP CAPACITY-BUILDING WORKSHOP 49

## Introducing More Rigor

Instead of ...	Consider producing stronger evidence ...
Logic model/theory	Correlational study
Correlational evidence	Quasi-experimental study
Quasi-experimental	Experimental study
Experimental study	A larger experimental study with varied populations

2016 NCEP/GEAR UP CAPACITY-BUILDING WORKSHOP 50

## Wrap Up – Resources Available



- National Evaluation i3 technical assistance, selected resources
  - [email NEi3\\_TA@abtassoc.com](mailto:NEi3_TA@abtassoc.com)
- Designing strong studies (webinar)
  - <http://ies.ed.gov/ncee/wwc/Multimedia.aspx?sid=18>
- Designing quasi-experiments (webinar, Slides)
  - <http://ies.ed.gov/ncee/wwc/event.aspx?sid=23>
  - [http://education.ufl.edu/educational-research/files/2015/04/ged\\_presentation\\_slides\\_030315.pdf](http://education.ufl.edu/educational-research/files/2015/04/ged_presentation_slides_030315.pdf)

Thank you for attending this workshop session.



## Quality of Project Evaluation (RFP)



- **Methods of evaluation are thorough, feasible, and appropriate** to the goals, objectives, and outcomes of the proposed project.
- Methods of evaluation include the **use of objective performance measures that are clearly related to the intended outcomes** of the project and will produce quantitative and qualitative data to the extent possible.
- Methods of evaluation will provide performance feedback and **permit periodic assessment of progress** toward achieving intended outcomes.
- Evaluation will provide **guidance about effective strategies** suitable for replication or testing in other settings.
- Methods of evaluation will, if well-implemented, **produce evidence of promise.**