

Identifying and Designing Studies that Meet *What Works Clearinghouse*TM Standards

Jonathan Jacobson, Ph.D.

jonathan.jacobson@ed.gov

Senior Research Scientist

National Center for Education Evaluation and Regional Assistance (NCEE)

Institute of Education Sciences (IES)

U.S. Department of Education

October 5, 2016

Overview

- Why Care About Evidence?
- *Strong Theory* and *Logic Models*
- *Evidence of Promise*
- The *What Works Clearinghouse™* (WWC)
- *Moderate Evidence of Effectiveness*
- Designing Evaluations to Meet **WWC Standards**

Why Care About Evidence?

- *“I like the **scientific spirit**—the holding off, the being sure but not too sure, the willingness to surrender ideas when the **evidence** is against them: this is ultimately fine—it always keeps the way beyond open—always gives life, thought, affection, the whole man, a chance to try over again after a mistake—after a wrong guess.” (Walt Whitman)*
- The U.S. Department of Education and its grantees are placing an increasing emphasis on **using** and **building empirical evidence** of the **effectiveness** of education programs, policies, and practices (**interventions**)

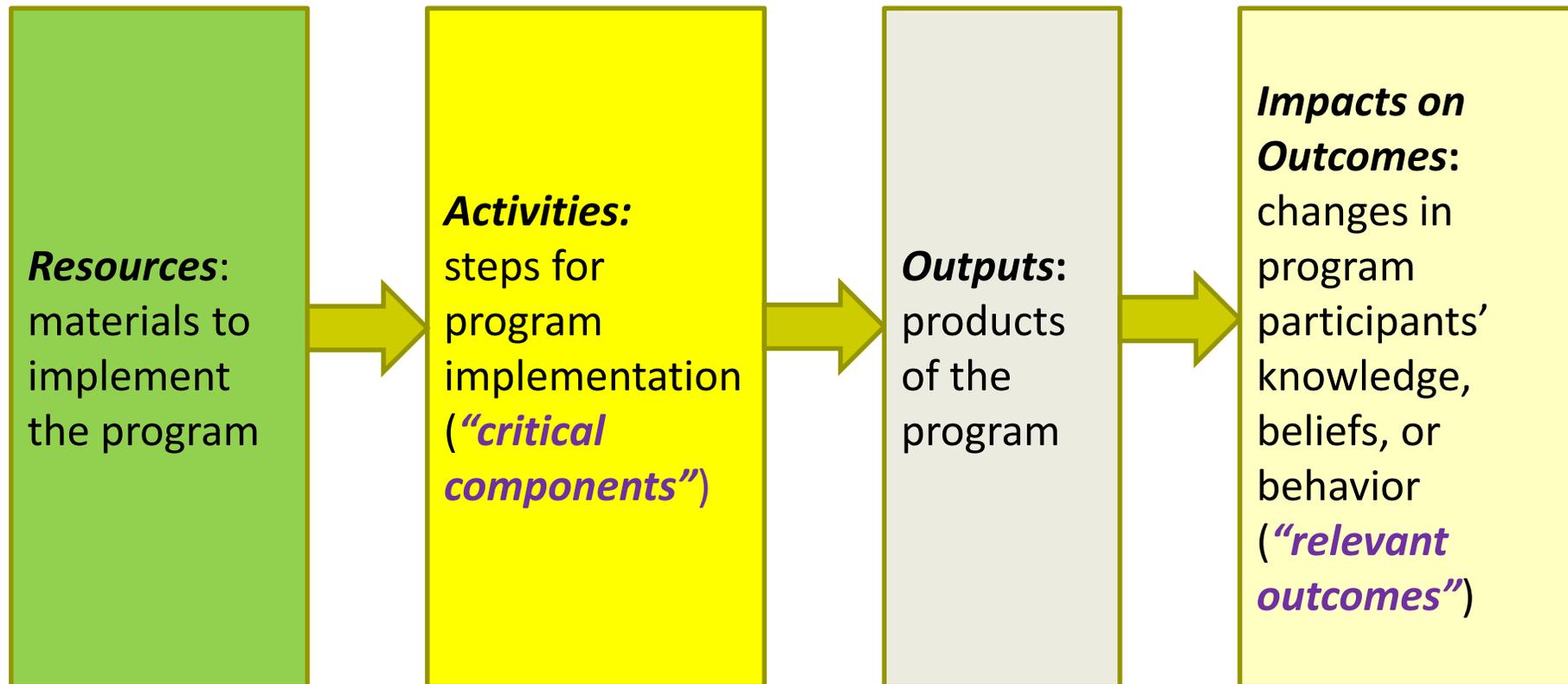
Strong Theory and Logic Models

(Education Department General Administrative Regulations,
Title 34 of Code of Federal Regulations, Part 77)

- ***Strong theory*** means “a rationale for the proposed process, product, strategy, or practice that includes a ***logic model***”
- A ***logic model*** (aka a ***theory of action***) means a well-specified conceptual framework that
 - identifies ***key components*** of the proposed process, product, strategy, or practice
 - describes the **relationships** among the key components and ***outcomes***

Components of a Program Logic Model

(from http://ies.ed.gov/ncee/edlabs/regions/pacific/pdf/REL_2014025.pdf)



Evidence of Promise

(Education Department General Administration Regulations, Title 34 of CFR, Part 77)

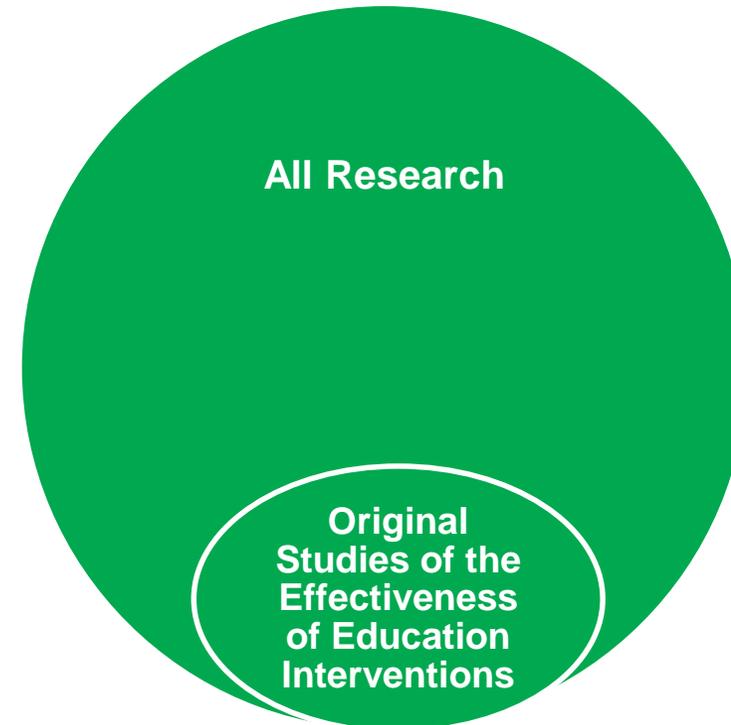
Evidence of Promise is “*empirical evidence* to support the theoretical linkage(s) between at least one *critical component* and at least one *relevant outcome* presented in the *logic model* for the proposed process, product, strategy, or practice.”

This must include at least one study that is either a—

1. *Correlational study* with statistical controls for selection bias;
2. *Quasi-experimental design (QED)* study that meets **What Works ClearinghouseTM** Evidence Standards with reservations; or
3. *Randomized controlled trial (RCT)* that meets **What Works ClearinghouseTM** Evidence Standards with *or* without reservations.

The What Works Clearinghouse™

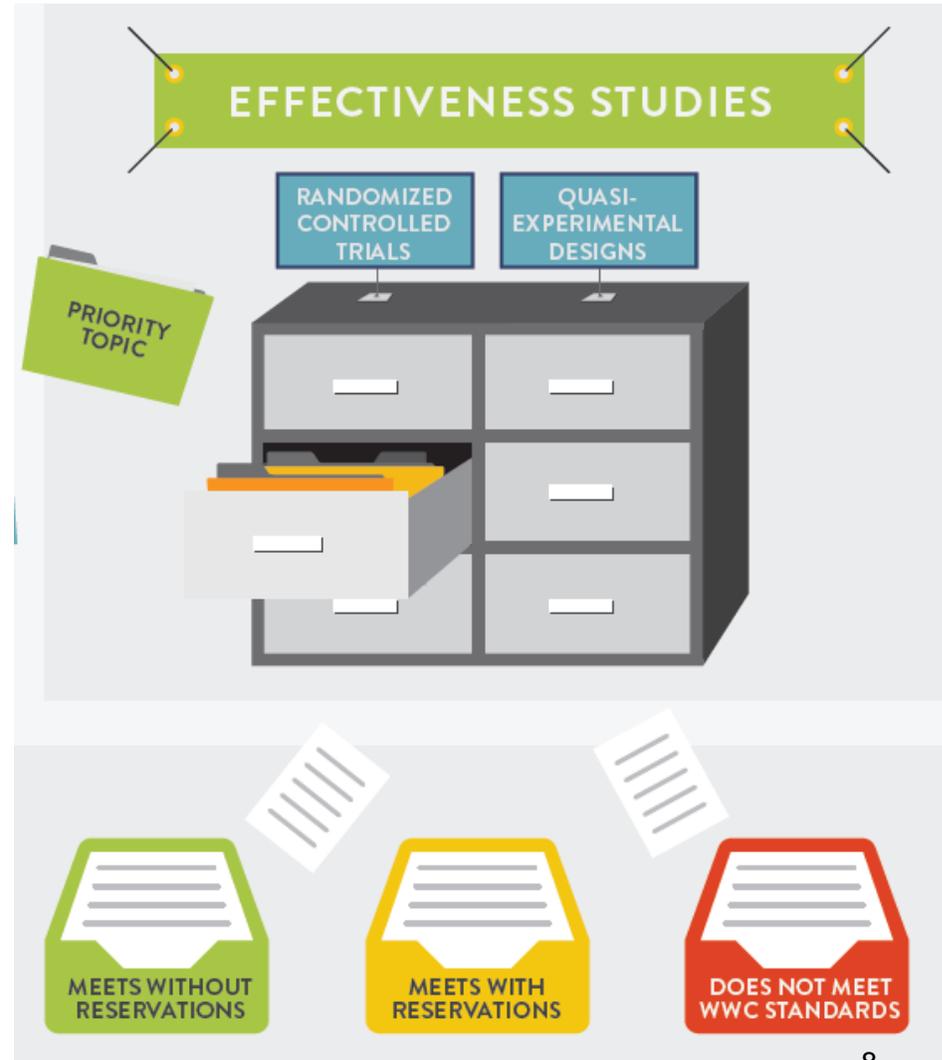
- Since 2002, the **WWC** has been an initiative of ED's Institute of Education Sciences (IES)
- The WWC reviews, rates, and summarizes **original** studies of the **effectiveness** of education **interventions**
- The WWC does **not** rate
 - ❖ Qualitative studies
 - ❖ Descriptive studies
 - ❖ Re-analysis or synthesis of others' data



Reviewed studies are documented on the WWC website; **findings** are reported from **studies that met WWC standards**

What Works Clearinghouse™ Standards

- Developed by panels of national experts for different types of designs for **effectiveness** studies
- Focus on **internal validity** of estimates (whether an estimated impact is valid or likely to be biased)
- Applied by teams of **certified reviewers** to give studies one of 3 ratings



Notes About WWC Evidence Ratings

1. WWC ratings are of study *findings*, not of *interventions*
2. The study as a whole receives the rating of the *highest-rated finding* reviewed by the WWC, *which may not be the most relevant finding for a particular project*
3. *Study ratings can change* when WWC standards change
4. The *sign, size, and statistical significance* of the estimated effect are reported by the WWC but do not affect the WWC study rating

New Tool for Finding Studies that Meet WWC Standards

Go to <http://ies.ed.gov/ncee/wwc/ReviewedStudies>, and select studies that:

- Meet WWC Standards,
- include at least one statistically significant and positive effect, and
- are in the topic area of interest

Evidence of Promise (continued)

To provide Evidence of Promise, a study must have “found a **statistically significant** or **substantively important*** favorable **association** between at least one *critical component* and one *relevant outcome* presented in the *logic model* for the proposed process, product, strategy, or practice.”

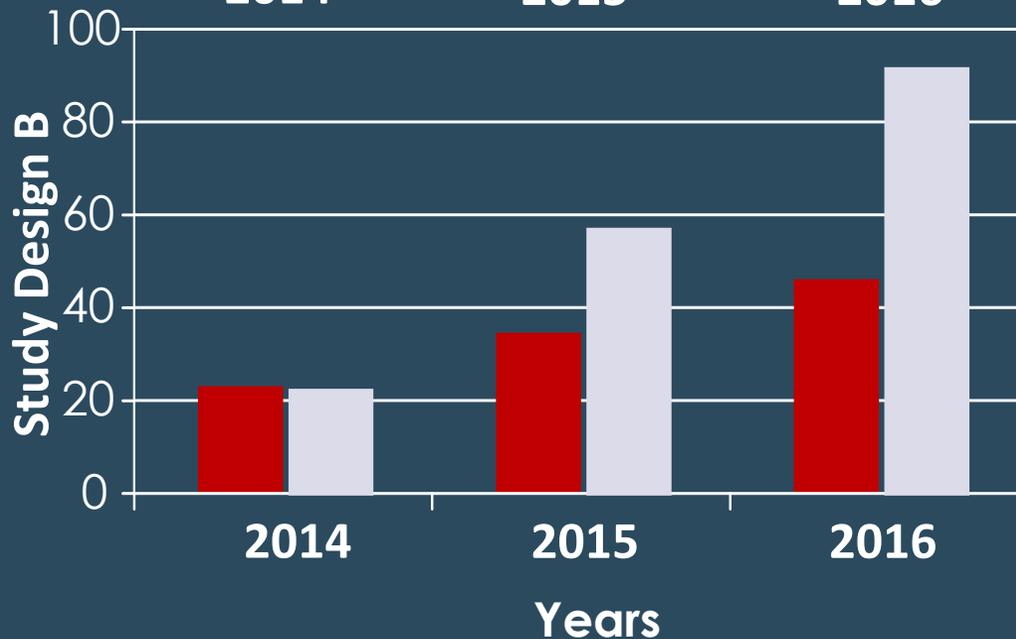
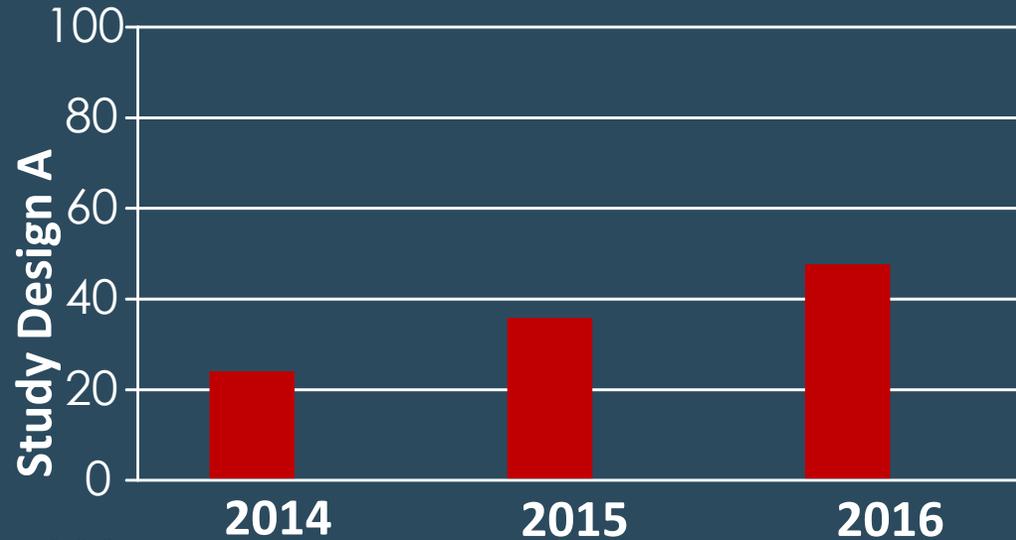
* = a **difference** of **0.25 standard deviations** or larger (for example, an impact of **+12.5 percentage points** if the comparison group mean is **50 percent**)

“Correlational study with statistical controls”

- A ***correlational study*** looks at the association between receipt of an ***intervention*** and an ***outcome*** of interest.
- ***Statistical controls for selection bias*** are the ***sampling or analytic methods*** study authors use to attempt to compare subjects ***similar*** except for the receipt of the ***intervention*** (“apples-to-apples” comparison)
- In general, a correlational study with statistical controls needs a ***comparison group*** and some effort to make it ***similar*** to the ***treatment group***

Why impact studies need comparison groups

13



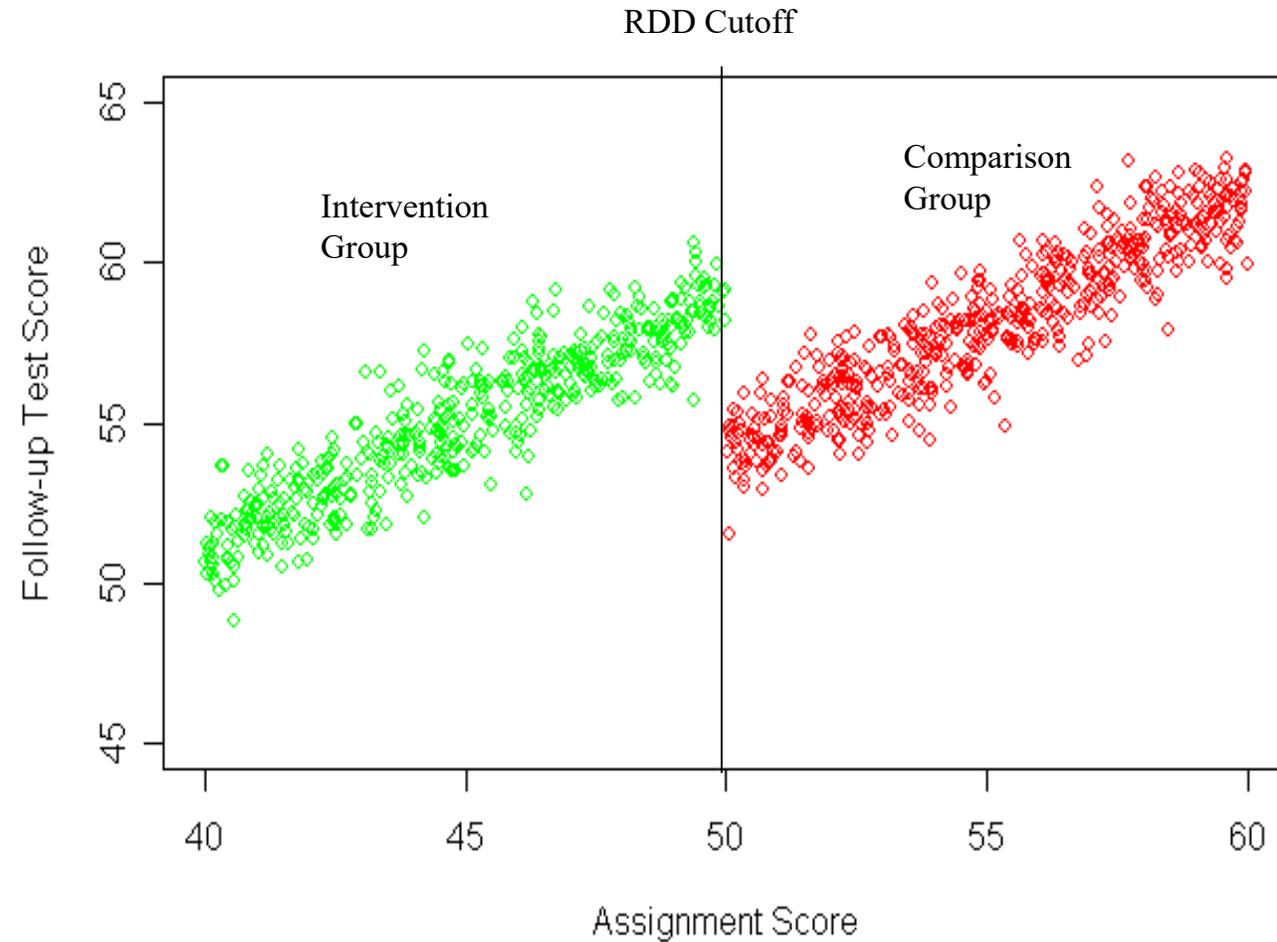
Suppose an intervention was first implemented in 2015 (2014 is baseline year, 2016 is year of follow-up data collection)

- Schools/students receiving intervention
- Comparison group (equivalent at baseline)

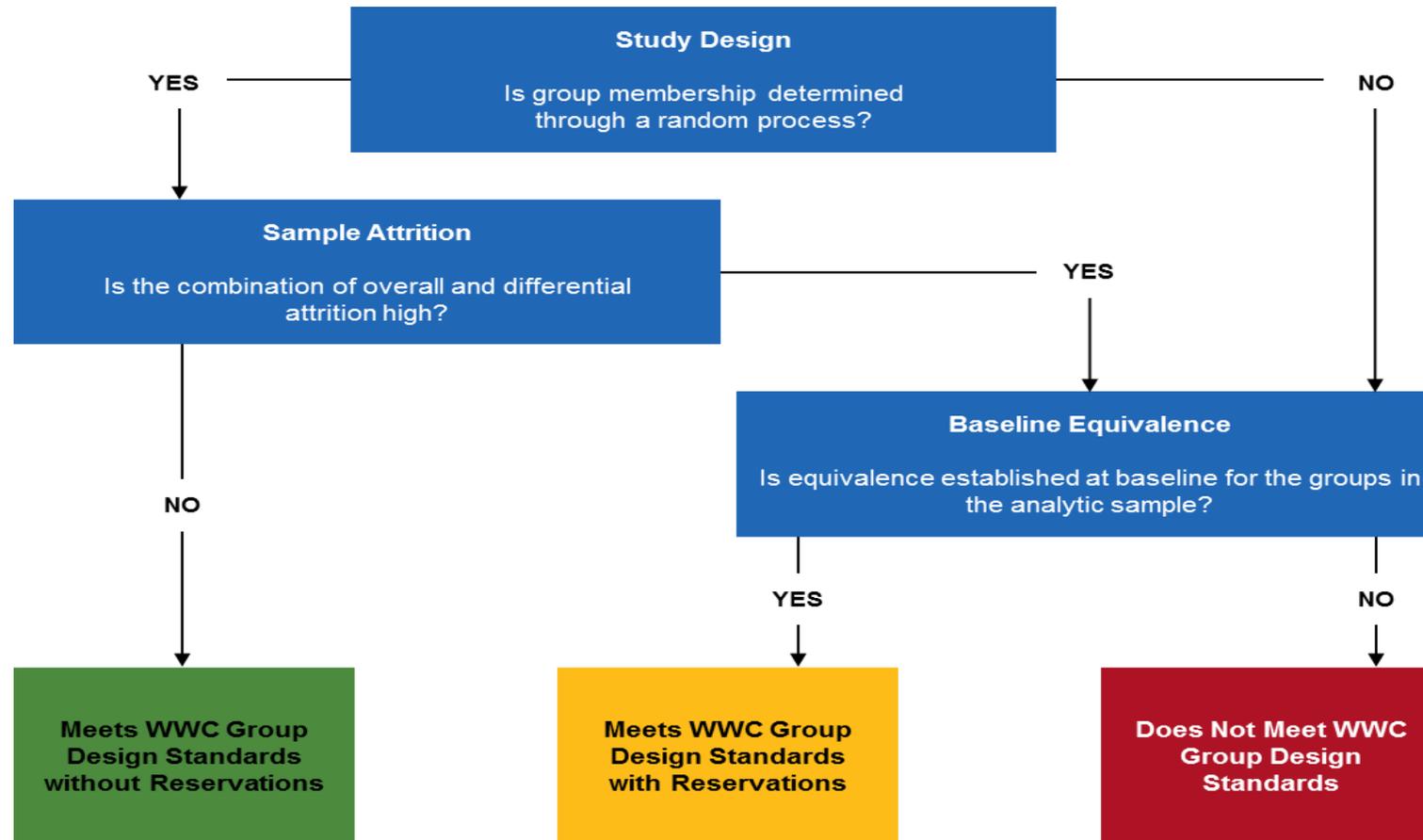
Ways to Form Comparison Groups for Impact Studies

1. Use a *lottery* to assign the *intervention (randomized controlled trial [RCT]—experimental design)*
2. Use an *index of need* and pick individuals on one side of the threshold to receive the *intervention (regression discontinuity design [RDD])*
3. Select the *treatment group* some other way, but use *pre-intervention characteristics* to compare similar groups of individuals in the impact analysis (*matched comparison design—quasi-experimental design [QED]*)

Estimation of Impact: Regression Discontinuity Design



WWC Ratings of a Group Design Study (RCT or QED)



A Study Providing *Moderate Evidence of Effectiveness...*

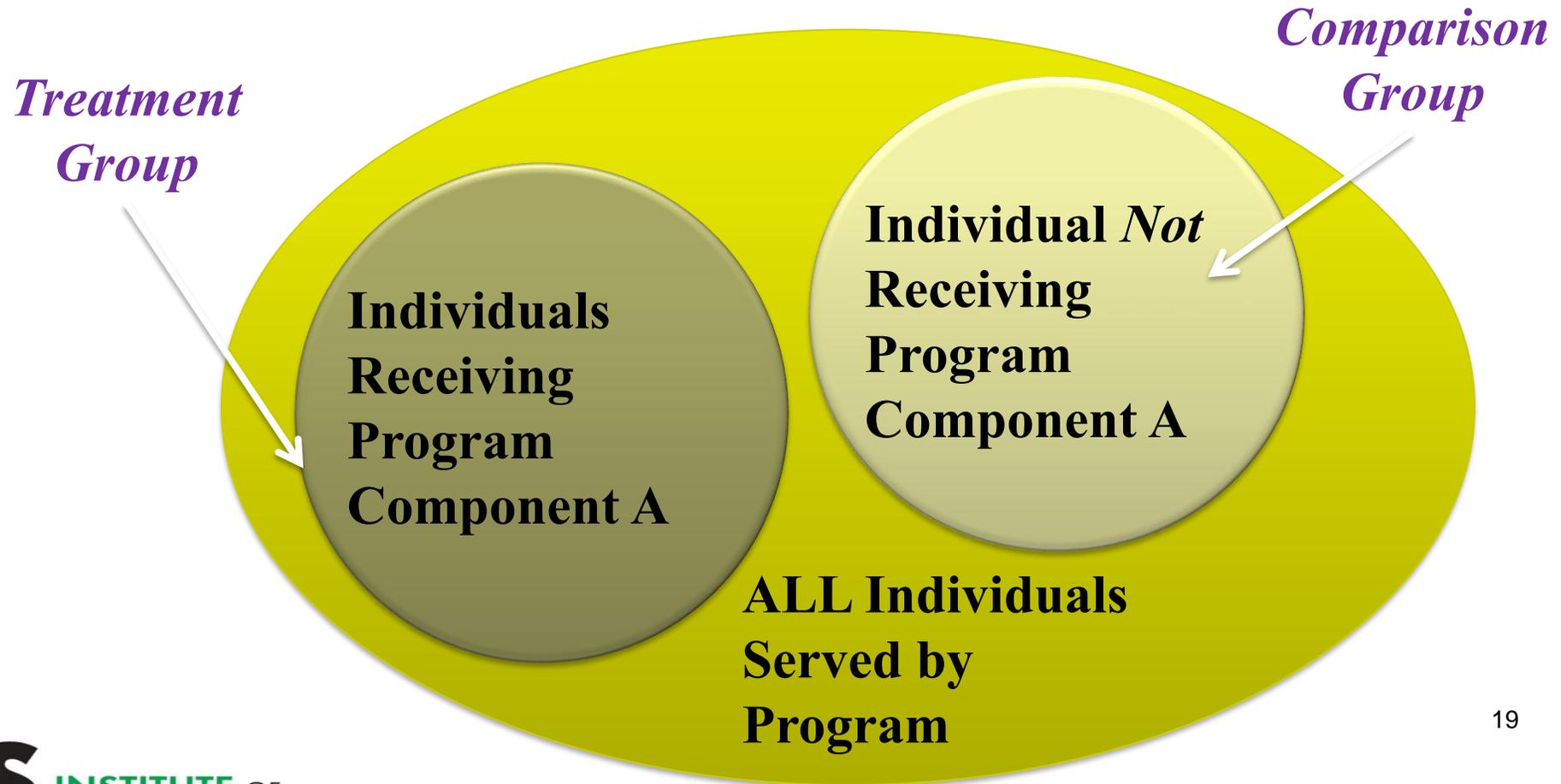
(Education Department General Administration Regulations, Title 34 of CFR, Part 77)

1. Is either (a) an **RCT** that **Meets WWC Standards Without Reservations**; or (b) a **QED** or **RCT** that **Meets WWC Standards With Reservations** and includes a **large sample** (350+ students, or 50+ clusters of at least 10 students/each) and a **multi-site sample** (> 1 LEA, locality, or state)
2. Has **overlap** with the population or settings proposed for the **intervention**
3. Shows a **statistically significant favorable impact** with no statistically significant and overriding unfavorable impacts in that study or other studies reviewed & reported by the **WWC**

Defining the Intervention for an Impact Evaluation

1. Which *intervention (critical component)* will we study that is supposed to affect a *relevant outcome* in our *logic model* (e.g., a key outcome we are already measuring for GPRA purposes)?
2. Is this intervention received by *all individuals* served by our project, or only by *some individuals*?
3. What services will be received by individuals in the *comparison group*, and do they offer a *service contrast* with the intervention?

Available Treatment and Comparison Groups among the Individuals Served by a Project



Resources for Designing Impact Evaluations to Meet WWC Standards

- Technical Assistance Materials for Conducting Rigorous Impact Evaluations <http://ies.ed.gov/ncee/projects/evaluationTA.asp>
- What Works Clearinghouse *Handbooks*
<http://ies.ed.gov/ncee/wwc/Handbooks>
- Webinars on Designing Strong Studies and QEDs
<http://ies.ed.gov/ncee/wwc/Multimedia/18>
<http://ies.ed.gov/ncee/wwc/Multimedia/23>
- Additional Resources on the Design of QEDs
<http://www.dir-online.com/wp-content/uploads/2015/11/Designing-and-Conducting-Strong-Quasi-Experiments-in-Education-Version-2.pdf>