



# Changes to the WWC 5.0 Standards & Procedures

---

Cristofer Price and Samuel Dastrup

EIR Annual Project Directors and Evaluators Technical Assistance Meeting

October 18, 19, 2022

# Asking Questions

---

- Participant microphones are automatically muted upon entry to limit background noise, unintentional feedback, or interference.
- During this session, you may ask the questions using the chat feature in Zoom.



- If your question is specific to your situation, or you need technical assistance, then the facilitator may respond to you directly in the chat panel.
- If your question is more general, then the facilitator will share your question with the presenter(s) so it can be answered at the appropriate time with the entire group.

# Agenda

---

- Topic-specific standards
- Definition of a study
- Bundled interventions
- Baseline equivalence requirements
- Difference-in-difference adjustment
- Independence of measures
- Compositional change and risk of bias due to joiners
- Domain level composites
- Effectiveness ratings

# Topic Area Protocols Play a Limited Role in WWC 5.0

---



- All WWC standards are now specified in the:
  - *Standards and Procedures Handbook* and
  - *Study Review Protocol*
- Topic area protocols will still be used for guidance regarding:
  - The literature search
  - How to identify and prioritize studies for inclusion in synthesis products
  - Intervention eligibility criteria
  - Eligible populations (e.g., grade ranges, locations)
  - Outcome eligibility criteria



# What This Means for You

---

- You are trying to design and implement an impact study that will meet WWC evidence standards
- Need to predict what standards are likely to be applied when your study is reviewed by the WWC
- This will still require familiarity with the:
  - Standards and Procedures Handbook (now version 5.0)
  - Study Review Protocol
  - Topic area protocols (especially regarding outcome eligibility)

**Design studies with backup plans in case your/our predictions are incorrect!**

# Agenda<sub>(2)</sub>

---

- Topic-specific standards
- **Definition of a study**
- Bundled interventions
- Baseline equivalence requirements
- Difference-in-difference adjustment
- Independence of measures
- Compositional change and risk of bias due to joiners
- Domain level composites
- Effectiveness ratings

# WWC 5.0: Findings from Independent Samples are Different Studies

## Version 4.1

Same study if findings have at least 3 of:

- Overlapping sample members
- Group formation procedures
- Data collection & analysis procedures
- Research team

## Version 5.0

Same study if findings have:

- Overlapping sample members
- Multiple sites in same manuscript

Different studies if findings have:

- Samples that do not share study participants
- Multiple sites in different manuscripts

# What This Means for You<sub>(2)</sub>

- How you report results might matter for multi-site or blocked studies
- Example: Study with sites A, B, and C
  - A and B are RCT and can meet with without reservations
  - C is QED and can meet with reservations

## Option 1

- Report A, B, and C in one manuscript →  
WWC will pool findings  
**Single** study will meet **with** reservations
- Tier 2 effectiveness rating

## Option 2

- Report A and B in manuscript I; C in manuscript II →  
WWC will review separately  
Study I meets **without**  
Study II meets **with** reservations
- Tier 1 effectiveness rating



# Agenda<sub>(3)</sub>

---

- Topic-specific standards
- Definition of a study
- **Bundled interventions**
- Baseline equivalence requirements
- Difference-in-difference adjustment
- Independence of measures
- Compositional change and risk of bias due to joiners
- Domain level composites
- Effectiveness ratings



# Bundled Interventions: No Longer Considered a Confounding Factor

## Confounding Factor in Version 4.1

- Intervention offered with a second intervention
- Can't isolate the effect of either intervention

## Not a Confounding Factor in Version 5.0

- Valid impact estimated for package of interventions



# What This Means for You<sup>(3)</sup>

---

- The WWC will review findings even if you evaluate a package of interventions that are bundled together
- Refer to the topic area synthesis protocol for information about eligible interventions for a given topic area

# Agenda<sub>(4)</sub>

---

- Topic-specific standards
- Definition of a study
- Bundled interventions
- **Baseline equivalence requirements**
- Difference-in-difference adjustment
- Independence of measures
- Compositional change and risk of bias due to joiners
- Domain level composites
- Effectiveness ratings

# Baseline Equivalence Requirements Under WWC 5.0



## Version 4.1

- All high-attrition RCTs (and RDDs) must establish baseline equivalence

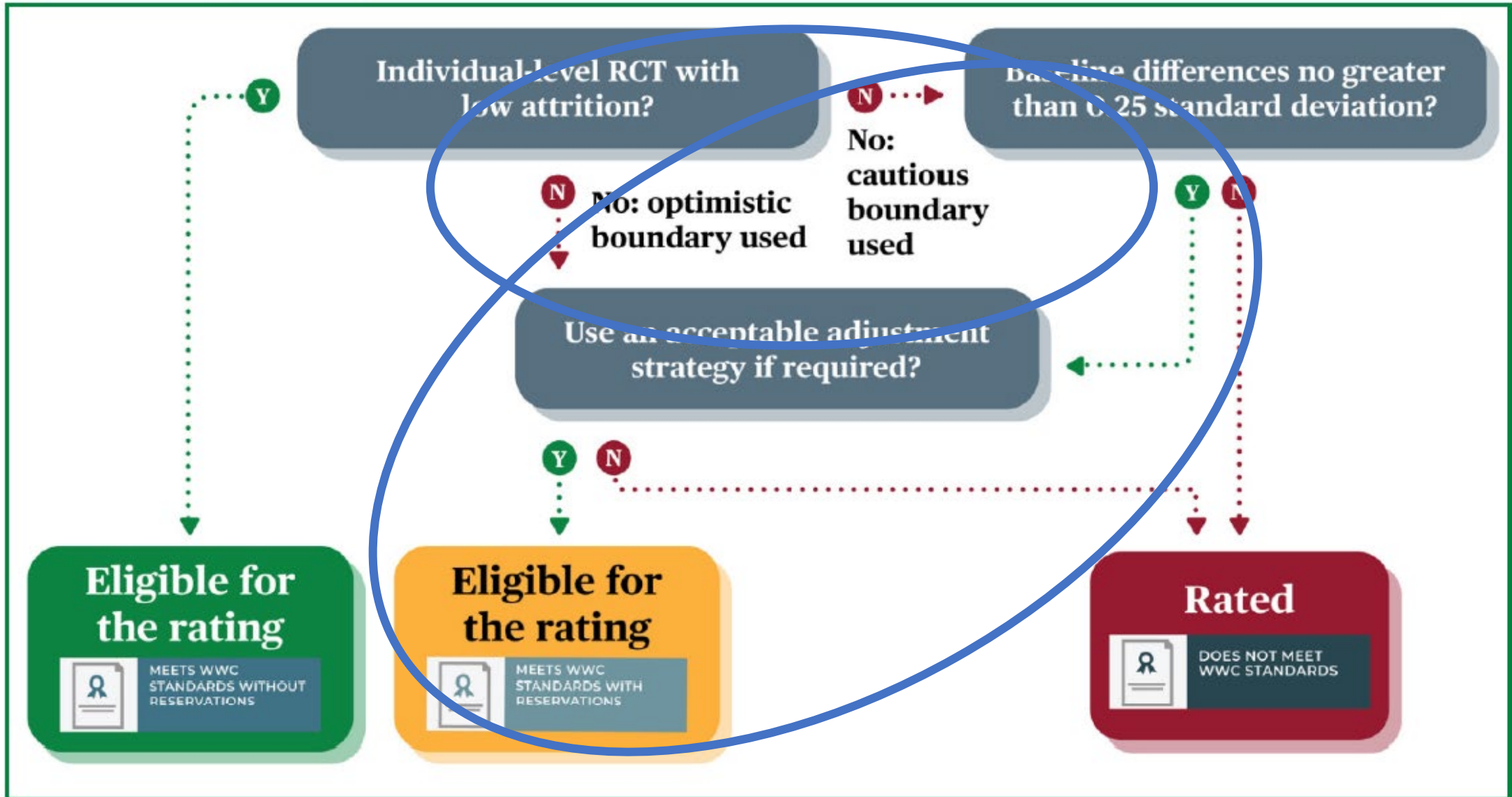
## Version 5.0

- High-attrition RCTs (and RDDs) **reviewed under optimistic attrition boundary** do *not* need to establish baseline equivalence
- Must adjust using acceptable strategy

Unchanged baseline equivalence requirements :

- Low-attrition RCTs (and RDDs)
- High-attrition RCTs (and RDDs) reviewed under *cautious* boundary
- All QEDs and compromised RCTs and RDDs

Figure 5. Ratings flowchart for individual-level assignment studies



# Baseline Equivalence Requirements Depend on the Attrition Boundary in WWC 5.0

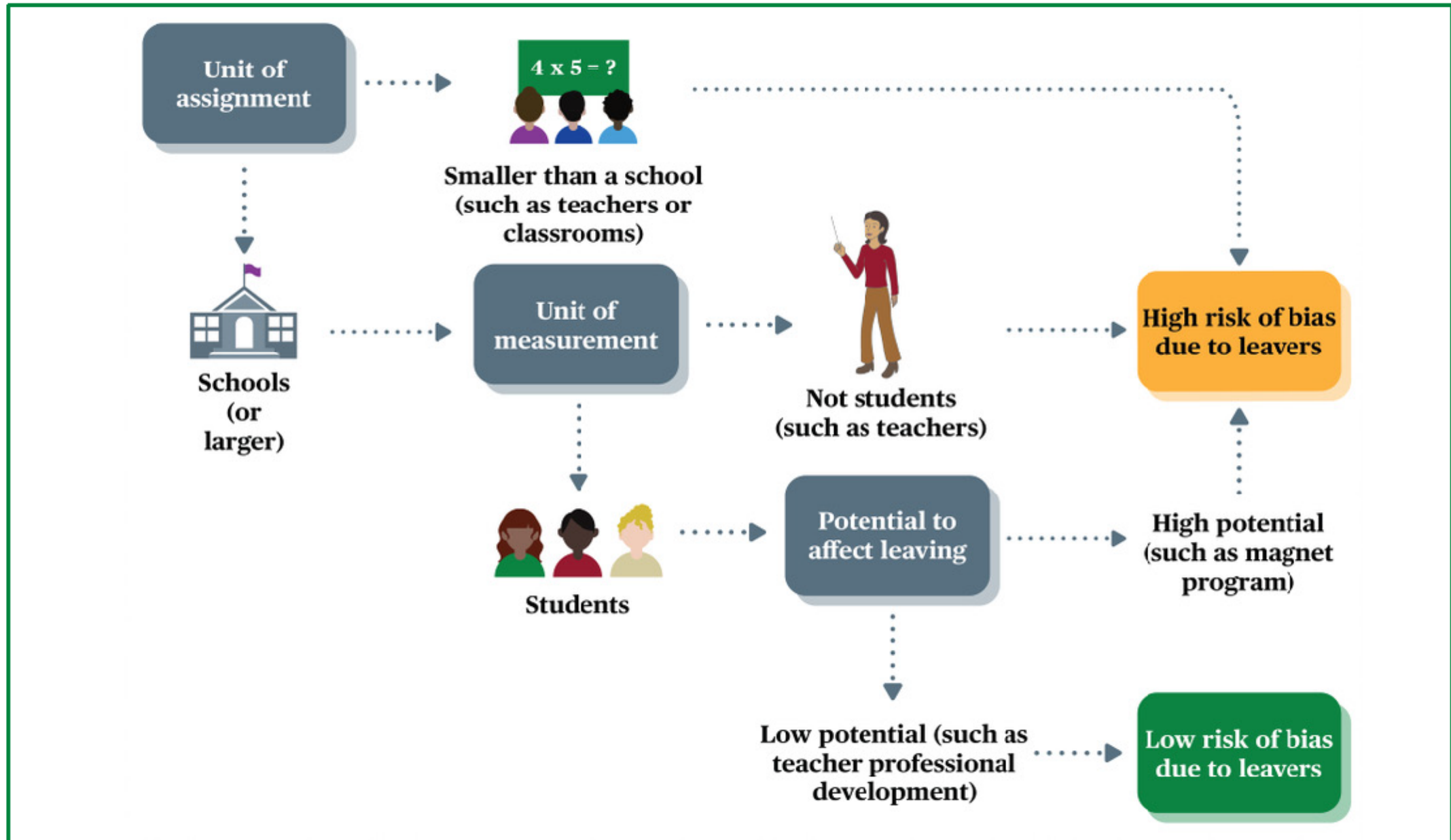


## Optimistic Boundary

- Attrition *unlikely* related to intervention
- Supplemental curriculum
- Targeted intervention during school hours
- PD program

## Cautious Boundary

- Attrition *likely* related to intervention (default)
- Dropout prevention
- School choice
- Elective/selective courses
- Outside of school hours





# What This Means for You<sup>(4)</sup>

---

- *Plan to demonstrate baseline equivalence, even if not required*
- It may be hard to predict whether a cautious or optimistic boundary will be used
- Describe intervention and comparison conditions with enough detail to allow for an optimistic boundary if appropriate
- Strategies to minimize attrition are still a good idea (best evidence)

# Agenda<sub>(5)</sub>

---

- Topic-specific standards
- Definition of a study
- Bundled interventions
- Baseline equivalence requirements
- **Difference-in-difference adjustment**
- Independence of measures
- Compositional change and risk of bias due to joiners
- Domain level composites
- Effectiveness ratings



# No Post-Hoc Difference-in-Difference Adjustments Under WWC 5.0

- WWC will no longer apply a difference-in-difference adjustment when study authors don't adequately adjust for baseline differences in their impact models
- If a study needs baseline adjustment according to the Standards, the study authors *must* be the ones to apply the adjustment (via author query)
  - If the author doesn't provide the adjustment, study will not meet standards.

# What This Means for You<sup>(5)</sup>

---

- For EIR evaluators, this is not a consequential change.
  - EIR evaluators already use acceptable methods for adjusting for baseline differences
- Typical approaches:
  - Baseline variable (i.e., the pre-test or its proxy) is a COVARIATE on the right-hand side of the impact model
  - Comparative Interrupted Time Series or Difference-in-Difference (But must show pretest-posttest correlation is 0.60 or greater)



# Agenda<sup>(6)</sup>

---

- Topic-specific standards
- Definition of a study
- Bundled interventions
- Baseline equivalence requirements
- Difference-in-difference adjustment
- **Independence of measures**
- Compositional change and risk of bias due to joiners
- Domain level composites
- Effectiveness ratings



# New Consideration In WWC 5.0: Independence of Outcome Measure

- Same four standards for eligible outcome measures:
  - Face validity
  - Reliability
  - Not overaligned
  - Consistent data collection procedures
- New consideration for ***some but not all*** outcome domains:
  - Independence of outcome measure
    - Outcome domains will be identified in a future version of the Study Review Protocol

# Independence of Outcome Measure Under WWC 5.0

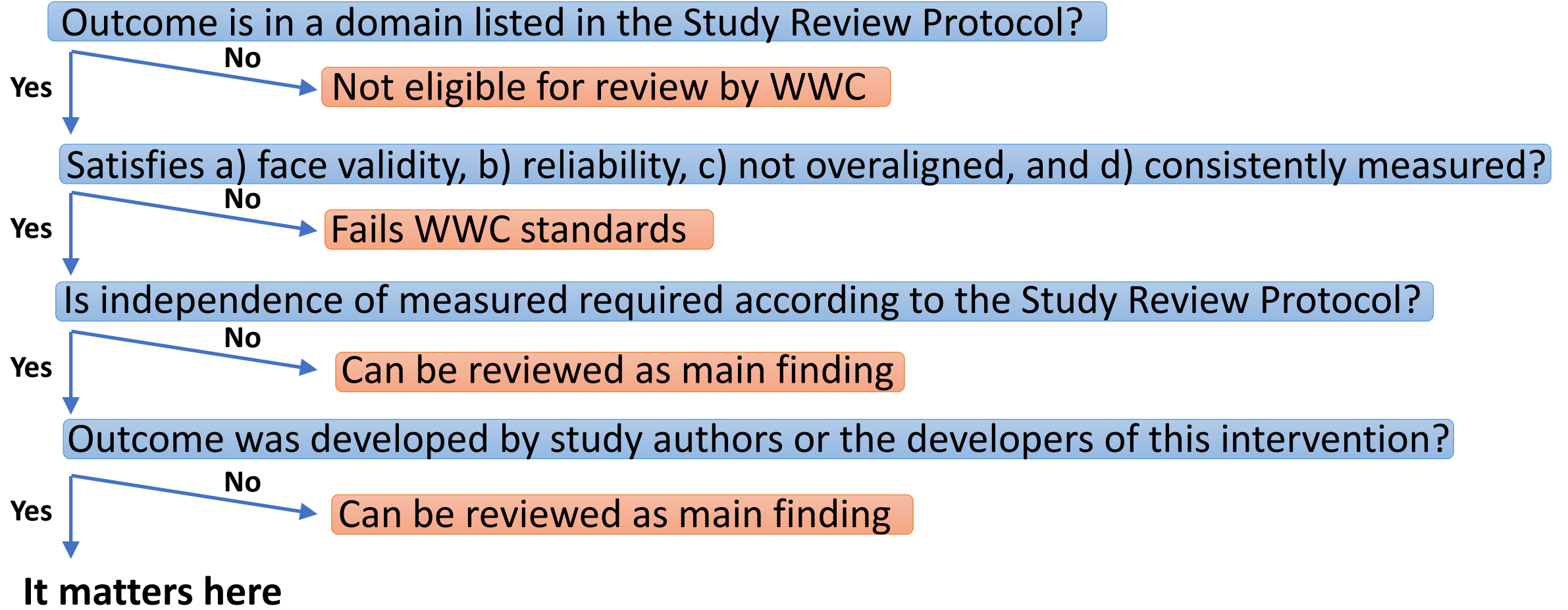
---



*“A measure will be considered non-independent if either it was developed by the study authors and is not in broader use, or if it was developed by the intervention’s developers”*

- If a measure was developed to be used with a particular intervention it will always be nonindependent if used as a measure for that same intervention

# Independence of Outcome Measure: When Does it Matter?







# What Happens if Your Outcome Measure is Deemed Nonindependent?

- Outcomes that meet standards AND deemed nonindependent in domains where independence matters will be reviewed as **supplemental findings**, not main findings
- Supplemental findings:
  - Can get “research ratings” (e.g., meets WWC evidence standards with or without reservations)
  - Do NOT contribute to effectiveness ratings (e.g., Tier 1 Strong Evidence, Tier 2 Moderate Evidence, etc.)

# What This Means for You<sup>(6)</sup>

---

- It's okay to use a measure that might be in a domain where independence matters AND deemed as nonindependent

**BUT**

- Don't place all your bets on that measure; Have at least one outcome measure that is CLEARLY independent



# Agenda<sub>(7)</sub>

---

- Topic-specific standards
- Definition of a study
- Bundled interventions
- Baseline equivalence requirements
- Difference-in-difference adjustment
- Independence of measures
- **Compositional change and risk of bias due to joiners**
- Domain level composites
- Effectiveness ratings

# Compositional Change and Joiners Under WWC 5.0

---



- We have a separate conference session devoted to this topic being presented by:
  - Anne Wolf and Eric Hedberg
    - Tuesday 10/18, 3:35 PM
    - Wednesday 10/19, 3:00 PM
- We recommend that you attend this session!

# Agenda<sup>(8)</sup>

---

- Topic-specific standards
- Definition of a study
- Bundled interventions
- Baseline equivalence requirements
- Difference-in-difference adjustment
- Independence of measures
- Compositional change and risk of bias due to joiners
- **Domain level composites**
- Effectiveness ratings

# No Multiple Comparisons Adjustments, Instead Composite Findings Created

## Version 4.1

When multiple main findings within an outcome domain WWC would:

- Conduct a Benjamini-Hochberg multiple comparison adjustment

## Version 5.0

When multiple main findings within an outcome domain WWC will:

- Combine the multiple findings into a single composite finding for the domain



# How Will the WWC Compute Composite Estimates under 5.0?

If you report multiple main findings that meet standards within an outcome domain the WWC will:

- Convert impact estimates and standard errors into effect size units (if you haven't already done so)
- Calculate the composite impact estimate as a simple mean of the effect sizes from the multiple findings:

$$\bar{g} = \frac{1}{K} \sum_{i=1}^K g_i$$

- Where  $g_i$  is the effect size from the  $i^{\text{th}}$  main finding

# What This Means for You<sup>(7)</sup>

---

- Just know that the WWC will compute a composite mean for domains with multiple main outcomes
- Consider if you'd prefer to:
  - Report only one outcome per domain; or
  - Create the composite yourself



# How Will the WWC Compute the Standard Error of the Composite Estimate?



The standard error for this domain-level average effect size will be calculated as follows:

$$SE[\bar{g}] = \frac{1}{K} \sqrt{\sum_{i=1}^K SE[g_i]^2 + \rho \sum_{i \neq j} SE[g_i]SE[g_j]}$$

where  $\rho$  is the average correlation among outcome measures, and  $SE[g_i]$  and  $SE[g_j]$  are the  $i^{\text{th}}$  and  $j^{\text{th}}$  effect size standard errors

# What This Means for You<sup>(8)</sup>

---

- Report the correlations between each pair of outcome variables in an appendix
  - WWC will use the study-reported correlations to calculate the domain average standard error
  - It will be less than the WWC-assumed correlation of 1
  - Your study will have greater power to detect effects



# How Will the P-Value of the Composite Estimate be Computed?

- A t-statistic will be calculated as the composite divided by its standard error:

$$t = \bar{g} / SE[\bar{g}]$$

- The *average degrees of freedom (df)* will be calculated as the mean of the degrees of freedom from your multiple impact estimates
- The p-value will be calculated using the t-distribution function in excel:

$$p = T.DIST.2T(t, df)$$



# What This Means for You<sup>(9)</sup>

---

- If you report impact estimates from multiple outcomes in the same domain then:
  - Report the degrees of freedom from each your impact estimates in an appendix
- If you do that, the reviewer won't have to make guesses about the degrees of freedom using potentially less favorable assumptions

# Agenda<sub>(9)</sub>

---

- Topic-specific standards
- Definition of a study
- Bundled interventions
- Baseline equivalence requirements
- Difference-in-difference adjustment
- Independence of measures
- Compositional change and risk of bias due to joiners
- Domain level composites
- **Effectiveness ratings**



# WWC's Research vs. Effectiveness Ratings

---

- **Research rating:** Is the evidence sufficiently rigorous?
  - Based on the rigor and execution of the study design
  - WWC rates individual findings, study inherits highest of the ratings
- **Effectiveness rating:** Did the intervention change outcomes?
  - Assessed for outcome domains that meet WWC standards
  - Based on composite of all main findings in each outcome domain
  - Aligned with ESSA Tiers of Evidence
  - Process varies for individual studies, intervention reports, and practice guides



# WWC Now Aligns Effectiveness Ratings with ESSA Tiers of Evidence

## Version 4.1

Characterize findings & extent of evidence *separately*

### Characterize finding(s) as:

- Positive effect
- Indeterminate effect
- Negative effect

### Extent of evidence:

- Medium to large
- Small

## Version 5.0

Characterize findings & extent of evidence in *combined effectiveness rating*

1. Strong evidence
  2. Moderate evidence
  3. Promising evidence
  4. Demonstrates a rationale (*practice guides only*)
- Uncertain Effects
  - Negative Effects

# Effectiveness Ratings

---

- Assessed by outcome domain using findings that meet standards (with or without reservations)
  - Main findings for study rating
  - Findings specified by WWC for synthesis products
- Effectiveness Rating depends on:
  - Research rating (i.e., with/without reservations)
  - Direction and statistical significance of findings
    - Including whether a negative impact in the same domain overrides a positive impact in a single study (intervention report or practice guide)
  - Number of sites
  - Number of individuals



# Criteria for Effectiveness Ratings: Individual Studies & Intervention Reports



TIER  
**1**  
STRONG

TIER  
**2**  
MODERATE

TIER  
**3**  
PROMISING

Positive and statistically significant effect on domain-level composite

More than 50% of findings rated *Meets WWC Standards Without Reservations*

More than 50% of findings rated *Meets WWC Standards With Reservations*

More than 50% of findings rated *Meets WWC Standards Without Reservations*

More than 50% *Meet WWC Standards* \_\_\_\_\_  
*Reservations*

With

Without

At least 350 unique individuals

At least 350 unique individuals

At least 20 unique individuals

Unique individuals less than

350

20

Multisite

Single site

No overriding negative effect in same domain



# Individual Studies & Intervention Reports

---

- **Uncertain Effects:**
  - The main effect composite is not statistically significant, or the significance is unknown or cannot be calculated; OR
  - The study has no main findings but at least one supplemental finding that meets WWC standards
- **Negative Effects:**
  - The main effect composite is statistically significant and negative

# What This Means for You<sub>(10)</sub>

- Effectiveness ratings are determined by the evidence rating, sample size, number of sites, and whether there are any overriding negative effects
- To get a Tier 1 or 2 effectiveness rating, main findings must:

	Grantee	Evaluator
<b>Design intervention that produces positive, statistically significant effects, and does not have overriding negative effects</b>	✓	
<b>Conduct study that meets evidence standards</b>		✓
<b>Include at least 350 students across at least two sites</b>	✓	✓

# Summary of Recommendations

---

- Try to anticipate what standards will be applied to your study
  - Know the handbook, study review protocol, and topic area protocols
- Have a backup plan in the event that study fails a standard
- Consider separate manuscripts if you have an RCT in some sites and a QED in others
- Don't worry about bundled interventions being ineligible for review
- Plan to demonstrate baseline equivalence
- Don't worry about changes to difference-in-difference adjustment procedures
- Have at least one measure that is clearly independent



# Summary of Recommendations<sup>(2)</sup>

---

- Be aware that the WWC will create a composite if you report more than one main finding in an outcome domain
  - Report the correlations between outcomes
  - Report the degrees of freedom from each analysis
- To get a Tier 1 or 2 effectiveness rating, main findings must:
  - Grantees: Design intervention that produces positive, statistically significant effects, and does not have overriding negative effects
  - Evaluators: Conduct study that meets evidence standards
  - Both: Include at least 350 students across at least two sites



Questions?

# Contact

---



EIR Evaluation TA Team

[EIREvalTA@AbtAssoc.com](mailto:EIREvalTA@AbtAssoc.com)



BOLD  
THINKERS  
DRIVING  
REAL-WORLD  
IMPACT

[www.abtassociates.com](http://www.abtassociates.com)