## WWC Cluster Design Standards

Research studies in which individuals are grouped within clusters have become more common in education research. This clustering can take a number of forms, including students within teachers, classrooms, or schools, and teachers or classrooms within schools. For ease of presentation, this document will discuss cluster designs including students within schools, though the standards presented here apply to all cluster designs.

Studies may involve random assignment or matching of schools but use information on students clustered within those schools to estimate impacts. In these studies, the observed effects of the program, practice, or policy can be influenced both by the direct effects of the intervention on students and by changes in the composition of students within the schools. This document presents criteria under which estimates of effects from cluster design studies *Meet WWC Group Design Standards Without Reservations* and the conditions under which they *Meet WWC Group Design Standards With Reservations*.

The process outlined in this document reviews the evidence of an intervention's direct effect on student-level outcomes (Steps 1–4). If a direct effect on student-level outcomes cannot be credibly demonstrated, the process reviews the evidence of an intervention's effect on cluster-level outcomes (Steps 5–7), where changes in the composition of students within the clusters may influence the observed effect. Each step involves addressing a question about the study's research design. The possible answers to each question lead to subsequent steps that should be taken as part of the review process. The review process is presented visually at the end of this document.

In the steps below, assessments of attrition and baseline equivalence will use standards described in the *Procedures and Standards Handbook* (version 3.0)*,* Sections III.B.2 and III.B.3.

### Screening criteria to determine whether a study has a cluster design

There are two key inclusion criteria for whether a study should be reviewed using the cluster design standards: (1) the unit of assignment in the study is a cluster, and (2) the data for the analysis are based on individuals within those clusters. If a study meets these two criteria, it should be considered a cluster design study and reviewed using the following steps. If the study does not meet either of the criteria, it should be reviewed as an individual-level design.

### I.  Process for reviewing evidence of an intervention's direct effect on individual-level outcomes (Steps 1–4)

The following four steps describe the review process to assess the credibility of the evidence in a study for understanding the direct effects of an intervention on individual-level outcomes.

### Step 1. Is the study a cluster RCT with low cluster-level attrition?

In order to receive the highest rating, the study must be a randomized controlled trial (RCT) with low cluster-level attrition.

- If the study is an RCT with low cluster-level attrition, continue to Step 2.

- If the study is an RCT with high or unknown cluster-level attrition, *or* is a quasi-experimental design (QED), continue to <u>Step 4</u>.

### Step 2. Is there a risk of bias due to individuals entering clusters?

In order to receive the highest rating, a cluster RCT must limit the risk of bias due to *joiners*, individuals entering the cluster after the time of random assignment or matching. If the study includes joiners in the analytic sample, the estimate of the direct effect of the intervention on student outcomes could be biased if the individuals who entered intervention clusters differ systematically from those who entered comparison clusters. This risk of bias may vary across substantive areas and interventions. Therefore, the review protocol for each area will define the risk of bias from joiners as high, moderate, or low.

> **High risk of bias associated with joiners.** Some reviews may include studies of programs or policies that are likely to affect enrollment or placement decisions, such as school turnaround interventions that close or combine schools, a policy that allows choice through the district, or the availability of charter schools. In these types of studies, joiners who enter intervention schools at any time after random assignment may be different than joiners who enter comparison schools because they may choose the school for a specific reason. For example, if high-performing students view the intervention schools as better suited for them and switch into those schools after the study begins, the observed effect may be biased if it was caused by differences in the types of joiners who entered the study. In this case, the potential for bias is high, because students choose the intervention schools for reasons specifically related to the intervention.

> **Moderate risk of bias associated with joiners.** Some reviews may include studies in which students who enter a school soon after the study begins (*early joiners*) are not likely to be a source of bias, but students entering later may be. For example, schools may be randomly assigned to implement a reading supplement or professional development program prior to or at the beginning of a school year. Some students may enter a school as the school year begins or shortly after because they have just moved to the neighborhood, which is common in many school settings. These early joiners are unlikely to have chosen the school for reasons related to the intervention, because the intervention is just beginning in the school and little may be known about it. Therefore, those early joiners may not differ from students who enter comparison schools early in the school year. However, students who enter schools later in the school year (*late joiners)* may be more likely to do so because of the intervention, and therefore differ from those who enter comparison schools later. The length of this initial period that differentiates between early or late joiners will be specified in the review protocol.

> **Low risk of bias associated with joiners.** Some reviews might focus on settings in which there is little or no risk of bias due to students who enter schools at any point after initial random assignment or matching. For example, interventions that have a very low profile, such as a change to recess programming, would not be expected to represent a significant draw for students. Protocols for those reviews may indicate that any students who enter schools after initial assignment (*both* early and late joiners) may be included in the analytic sample without risk of bias.

- If the study either (1) excludes from the analytic sample joiners who enter after random assignment or (2) only includes in the analytic sample joiners who enter during the period(s) determined to not be associated with a risk of bias, move to Step 3.

- If the study includes in the analytic sample any joiners who entered during a period determined to be associated with a risk of bias, it must demonstrate baseline equivalence of the analytic sample in Step 4.

## Step 3. Is there a risk of bias due to non-response of individuals?

In order to receive the highest rating, a cluster RCT with a limited risk of bias due to joiners must also have low individual-level attrition. Attrition is the difference between the sample present in a reference population (described below) and those present at the time the outcome is assessed (i.e., the analytic sample). The reference population—the baseline sample from which attrition is measured—can differ for studies with high, moderate, or low risk of bias associated with joiners.

When the **risk of bias from joiners is high,** the only relevant reference population is (1) the sample of students in non-attriting clusters who were subject to random assignment (i.e., students in clusters at or before the time of random assignment).

When the **risk of bias from joiners is moderate,** in addition to reference population (1), the reference population can also be (2) the sample of students observed during an initial period (e.g., the rosters of students obtained from a school early in the school year) in non-attriting clusters.

When the **risk of bias from joiners is low,** in addition to reference populations (1) and (2), the reference population can also be (3) the sample of students at follow-up (e.g., the number of students enrolled in study schools on the day the posttest was given) in non-attriting clusters.

When the risk of bias from joiners is moderate or low, multiple possible reference populations are available. Operationally, if a study provides information for multiple reference populations, and attrition can be assessed relative to each population, the WWC will prioritize the earliest available point in time for calculations of individual-level attrition.

- If the study has low levels of individual-level attrition, it is eligible to *Meet WWC Group Design Standards Without Reservations*.

- If the study has high levels of individual-level attrition, it must demonstrate baseline equivalence of the analytic sample in Step 4.

## Step 4. Is baseline equivalence established for the analytic sample of individuals?

For cluster RCTs with a risk of bias from high cluster-level attrition, high individual-level attrition, or from unallowable joiners, and for cluster QEDs, baseline equivalence of the analytic sample must be established for the study to *Meet WWC Group Design Standards With Reservations*.

- If the study shows baseline equivalence of the analytic sample, then it is eligible to *Meet WWC Group Design Standards With Reservations*.

- If the study does not show baseline equivalence of the analytic sample, then it will be reviewed to determine if it can provide credible evidence of the intervention's effect on cluster-level outcomes. This review process is described in the following section, beginning with Step 5.

## II. Process for reviewing evidence of an intervention's effect on cluster-level outcomes (Steps 5–7)

The following three steps describe the review process to assess the credibility of the evidence in a study for understanding the effects of an intervention on cluster-level outcomes. In these studies, the observed impact estimate potentially represents a combination of (1) the direct effect of the intervention on individual-level outcomes and (2) a composition effect due to different types of individuals entering intervention and comparison clusters. Therefore, evidence reviewed in this section is only eligible to *Meet WWC Group Design Standards With Reservations*.

### Step 5. Is the analytic sample of individuals representative of the clusters?

The attrition standard for group design studies will also be applied to cluster RCTs to assess the representativeness in making inferences for the cluster. The numerator for the attrition calculation will be the number of individuals contributing to the impact estimate. The denominator for the attrition calculation will be the number of individuals in clusters (contributing to the impact analysis) at the time of the follow-up period. If the study has poor response rates at follow-up, or sufficiently differential response rates among individuals in the intervention and comparison clusters, the observed impact would not credibly estimate the effect of the intervention on cluster-level outcomes.

- If the study has low individual-level attrition for this representativeness assessment, move to Step 6.

- If the study has high or unknown individual-level attrition for this representativeness assessment, it *Does Not Meet WWC Group Design Standards.*

### Step 6. Is there a risk of bias due to non-response of clusters?

NOTE: This is the same assessment of cluster attrition from Step 1.

- If the study is an RCT with low levels of cluster attrition, it is eligible to *Meet WWC Group Design Standards With Reservations*.

- If the study is an RCT with high or unknown cluster attrition, or is a QED, then move to Step 7.

### Step 7. Is baseline equivalence established for the analytic sample of clusters?

For cluster RCTs with high or unknown attrition and cluster QEDs, baseline equivalence of the clusters contributing to the analytic sample must be established for the study to *Meet WWC Group Design Standards With Reservations*.

Each review protocol will indicate the characteristics for establishing baseline equivalence of clusters. Examples of characteristics may include student achievement levels, grade levels, demographics of teachers or students in schools, and school setting (urban or rural). The review protocol will also determine the parameters for establishing equivalence. For example, a protocol may allow equivalence to be established for an analytic sample of fourth graders in 2015 using an earlier cohort of students that does not largely overlap with the analytic sample (e.g., fourth graders in 2014) or an earlier assessment of the same cohort in the analytic sample (e.g., the same cohort in third grade in 2014).

Additionally, the students with baseline data must be representative of the clusters contributing to the impact analysis, assessed by comparing the number of students with baseline data relative to the number of students in the clusters at the time of the baseline equivalence assessment. This representativeness assessment for baseline data is analogous to the representativeness assessment for follow-up data described in Step 5. There is some flexibility to allow the reference population for this baseline equivalence assessment to be any point in time during the baseline year.

- If the study demonstrates baseline equivalence of the clusters in the analytic sample, and there is low attrition of individuals relative to the baseline population, the study is eligible to *Meet WWC Group Design Standards With Reservations*.

- If the study does not demonstrate baseline equivalence of the clusters in the analytic sample, or there is high attrition of individuals, the study *Does Not Meet WWC Group Design Standards*.

**Review process for cluster design studies under version 4.0 standards**