

Answers to Technical Questions from *Understanding the Updated WWC Procedures and Standards*

Question: What is the rationale for permitting researchers to choose their denominator (e.g., spring rather than fall semester) if they demonstrate low joiner risk?

Response: The cluster-level assignment standards aim to provide reasonable flexibility to study authors in defining the reference sample for assessing individual-level non-response. For example, flexibility may be needed because in many cluster-level assignment studies it is not possible to identify the individuals who were within clusters at the time of random assignment, such as when teachers are assigned to conditions in the summer before the start of a school year, before classroom assignments have been determined.

The WWC will calculate individual-level non-response for an analytic sample that includes joiners, provided the joiners satisfy the requirement that they do not pose a risk of bias, according to the applicable review protocol. As stated on page 26 of the *WWC Standards Handbook*, an acceptable reference sample must be defined at a point in time after all joiners included in the analytic sample had already joined clusters. Consequently, researchers are constrained by the need for the reference sample to include all of the joiners in the analytic sample. In the example suggested in the question, where researchers included individuals who joined clusters during the spring semester in the analytic sample, the only acceptable reference samples are those based on enrollment late in the school year. Those based on enrollment during the fall semester would not be acceptable.

Also, as noted on page 27 of the *WWC Standards Handbook*, if a study provides information on multiple acceptable reference samples for assessing individual-level non-response, the WWC will base its calculations on the earliest sample.

Question: In a cluster RCT in which participating school districts require active consent to participate in research activities, and the student consent process takes place after the beginning of the school year, should student non-response attrition calculations use as the denominator: a) all students who received consent forms, in the denominator or, b) all students with consent?

Response: Some sample exclusions can be excluded from the denominator of attrition, non-response, and representativeness calculations in cluster randomized controlled trials and from representativeness calculations in quasi-experimental designs, as discussed on page 31 of the *WWC Standards Handbook*. In particular, they can be excluded when the exclusion was based on characteristics that were clearly determined prior to the introduction of the intervention (i.e., before students knew their assigned conditions) and applied consistently across the intervention and comparison groups. Therefore, if consent was collected prior to the introduction of the intervention and performed in the same way in both groups, the non-consenting students may be excluded from the denominator for these calculations.

Question: Are WWC ratings aligned with the ESSA evidence levels? For example, if a study *Meets WWC Standards Without Reservations*, is it equivalent to the ESSA strong evidence?

Response: The Every Student Succeeds Act of 2015 (ESSA) included definitions of “strong evidence” from at least one “well-designed and well-implemented experimental study” and

“moderate evidence” from at least one “well-designed and well-implemented quasi-experimental study.” In both instances this evidence for an “activity, strategy, or intervention” must include “a statistically significant effect on improving student outcomes or other relevant outcomes” (see <https://www2.ed.gov/policy/elsec/leg/essa/legislation/title-viii.pdf>). While the legislation itself did not mention the WWC, the [non-regulatory guidance](#) released by ED in September 2016 and the [evidence definitions](#) announced in July 2017 that apply to ED direct grant programs both refer to studies needing to meet WWC standards in order to provide “moderate evidence” or “strong evidence.”

A study’s WWC rating is not the only factor in the definitions of strong evidence and moderate evidence under ESSA. A study can meet WWC standards and not include statistically significant findings. In addition, there are large sample and multi-site sample requirements specified by ED in order for a study or group of studies to provide strong evidence or moderate evidence. Consequently, a rating of *Meets WWC Standards With Reservations* or *Meets WWC Standards Without Reservations* is not sufficient, by itself, to establish that a study provides moderate or strong evidence of effectiveness.

Question: Clarification on baseline equivalency: For a QED using matching and computing simple annual growth measures across multiple years, is it necessary to show baseline equivalency at every year’s growth, or is it sufficient to show baseline equivalency in the initial match year only?

Response: In an individual-level assignment quasi-experimental design study, baseline equivalence must be assessed on the exact analytic sample of individuals. So, if the individuals in the analytic sample change over time, then baseline equivalence must be assessed separately in each year. In general, the baseline period may be kept the same for all outcome time periods.

However, for some analyses of growth measures, the baseline period used to assess baseline equivalence cannot be earlier than the baseline period for calculating the growth measure. Consider two types of growth analyses with outcomes taken from the spring of multiple school years. In the first, growth is always measured relative to the same baseline period. In the second, growth is measured relative to a fall score in the same school year as the outcome measure. When growth is always measured against the same baseline period, then it is acceptable to use the same (or later) period for assessing baseline equivalence for all outcome time periods. In the second type of growth analysis, the earliest baseline period may not be acceptable for assessing baseline equivalence for outcome measures in later periods because the fall measures used to calculate growth may have been affected by group status. (As described on page 35 of the *WWC Standards Handbook*, a covariate that was assessed or obtained after baseline is considered endogenous if it was potentially influenced by group status. An analysis with an endogenous covariate will be rated *Does Not Meet WWC Group Design Standards*.) Instead, such an analysis would need to satisfy the baseline equivalence requirement using baseline data from the period of the measure used to calculate growth (e.g., the fall of the school year of the outcome measure).

Question: Assume that baseline and/or outcome measures were imputed. When determining the correlation between outcome and baseline measures for alternative statistical adjustment requirements, would that be based on non-missing measures or imputed measures?

Response: In general, the correlation used to assess whether fixed effects, gain scores, or difference-in-differences adjustments are acceptable should be based only on observed data, excluding observations with imputed data. However, as noted on page 15 of *the WWC Standards Handbook*, topic areas have some discretion here, so that some topic areas may allow assessment based on data that includes imputed values. On page 42 of the *Handbook*, the missing data standards require that the same correlation, which is also used to assess concerns related to missing or imputed data, must be based only on observed data.

Question: Is it required that all or some of the project team and/or evaluators have WWC certification?

Response: The WWC does not set any requirements for certification of members of project teams or evaluators. We recommend that you check with your funder or project officer for any requirements they might have.

The WWC offers three levels of certification through the WWC training programs.

The first level is to receive a certification of completion for viewing all of the videos in the [online training program](#). The online training also prepares those who are interested in pursuing the optional group design certification process (the second option described below). This training is ideal for those interested in assessing the credibility of research evidence on the effectiveness of educational interventions and interpreting research findings. This may include people trying to make evidence-based decisions in educational settings or researchers who are developing rigorous studies to meet WWC group design standards. If you are unsure if this option is right for you, we recommend viewing the Introduction module for this training series first.

The second level is to become certified in the WWC group design standards. To receive this certification, you must successfully complete a certification exam that tests your knowledge of the standards covered in the online training modules. Details on this will be coming soon on the WWC website.

The third level is to become a certified reviewer for the WWC. After the in-person training, participants are expected to successfully complete a reviewer certification, which includes an exam. The WWC certification process is a rigorous program designed for experienced participants that possess strong design and analytic skills. Participants must be able to understand technical concepts about study design, calculating effect sizes, and determining the statistical significance of findings. To receive this certification, you must be certified in the WWC group design standards first. Becoming a certified reviewer requires successful completion of additional tasks. Those who successfully complete this level of certification can be included in the list of certified reviewers on the WWC website, and the WWC or the Institute of Education Sciences may contact them to assist with review efforts.