WWC STANDARDS Brief

The What Works Clearinghouse (WWC) is an initiative of the U.S. Department of Education's Institute of Education Sciences. The WWC evaluates research studies that look at the effectiveness of education programs, policies, and practices, which the WWC calls "interventions." WWC Standards Briefs explain the rules the WWC uses to assess the quality of studies. For more information, visit the WWC's webpage at http://whatworks.ed.gov.

Baseline Equivalence Standard

What is baseline equivalence?

The WWC uses the term "baseline equivalence" when determining if the intervention group (those that received the intervention of interest) and the comparison group (those that did not receive the intervention) had characteristics that were similar enough ("equivalent") prior to or at the start of the study (at "**baseline**").

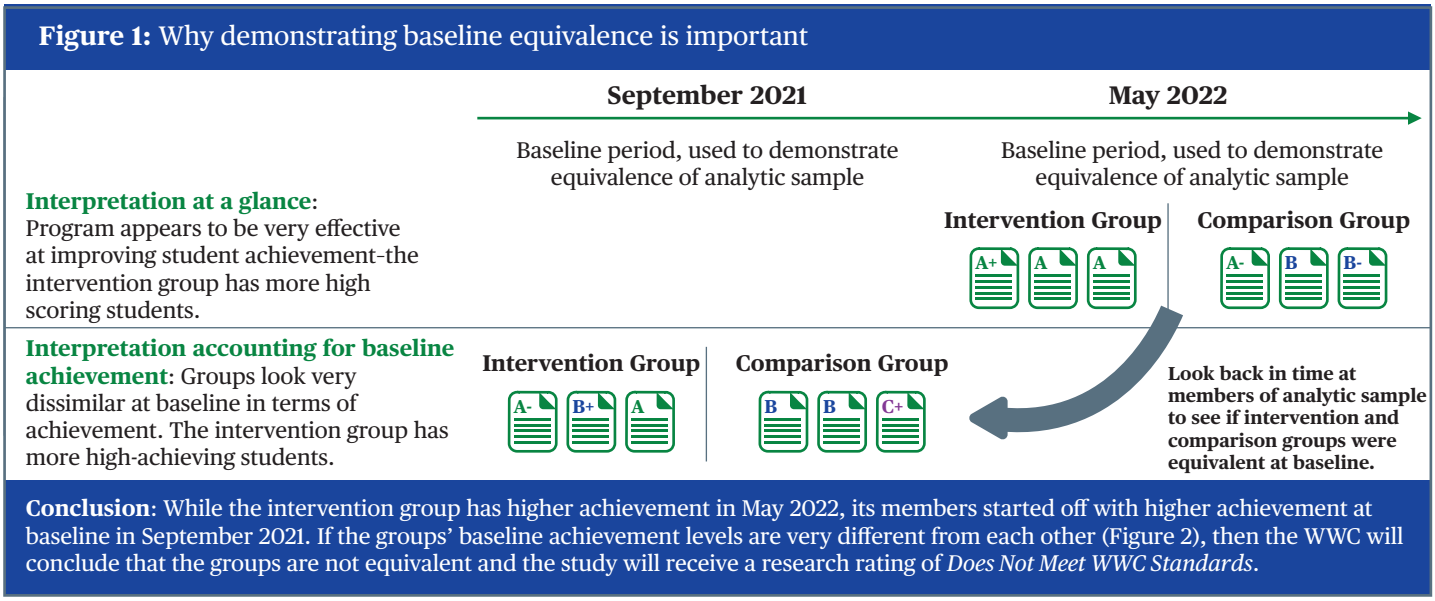
Why does baseline equivalence matter?

When two groups are similar at the start of a study (baseline) and, after that, the only difference between the groups is that one receives the intervention and the other does not, it is reasonable to conclude that any differences in the outcomes that are measured at the end of a study (**follow-up**) are caused by the intervention. However, if the two groups are different at baseline on key characteristics that could influence the outcomes, the effect found at the end of the study might be due to the differences that already existed at the beginning.

Demonstrating baseline equivalence is important in studies that did not assign participants randomly to the intervention and comparison groups. It also is important in random assignment studies if random assignment was compromised and in some random assignment studies with high attrition.[1]

In the example in Figure 1, students in the intervention group had higher academic achievement than those in the comparison group during the follow-up period, suggesting a positive program impact. However, we need to look back in time and examine the baseline achievement for the sample of individuals analyzed at follow-up (the **analytic sample**).

Figure 1: Why demonstrating baseline equivalence is important



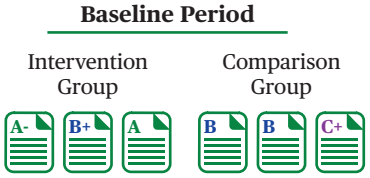
[1] Attrition refers to instances in which units (such as students, teachers, or principals) that were assigned to intervention or comparison conditions leave the study or are otherwise unavailable for outcome measurement. See Chapter 3, Step 3 in the WWC Procedures and Standards Handbook. See the WWC Standards Brief for Attrition for more information on this topic.

The groups started out at very different achievement levels: the intervention group had more high-achieving individuals than the comparison group had at baseline. Therefore, the observed program impact may be biased: some of the differences in outcomes may result from having different types of students across intervention and comparison groups.

How does the WWC determine baseline equivalence?

The WWC Procedures and Standards Handbook, Version 5.0, specifies the characteristics on which equivalence must be established at baseline. For academic outcomes, baseline equivalence is often established using an achievement pretest. Some outcomes, such as high school graduation, do not have a pretest. For these outcomes, the WWC requires baseline equivalence on several characteristics, most often achievement plus two related characteristics (e.g., demographics, grade level).

Figure 2: Baseline equivalence calculation



Group	Baseline Mean	Baseline Standard Deviation
Intervention	90	10
Comparison	82	10

Difference in means ▶ 90 - 82 = 8

Pooled standard deviation ▶ 10

Baseline Equivalence Calculation

$$\textit{Effect size} = \frac{\textit{Difference in means}}{\textit{Pooled standard deviation}} = \frac{8}{10} = .80$$

Decision

Does not satisfy baseline equivalence

The WWC uses a standardized mean difference called an **effect size** to determine whether there is baseline equivalence between the intervention and comparison groups. The effect size is calculated as the difference between the intervention group mean and the comparison group mean, divided by the pooled standard deviation.[2] Figure 2 illustrates this calculation. In this example, the intervention group had higher average achievement at baseline than the comparison group. The WWC then determines whether the groups are similar by comparing the effect size for each baseline characteristic against the WWC standard for baseline equivalence in Figure 3.

The WWC considers intervention and comparison groups equivalent at baseline when the corresponding effect sizes are 0.05 or less in **absolute value**. When an effect size for a baseline characteristic is between 0.05 and 0.25 in absolute value, the WWC requires a **statistical adjustment**.[3] If the effect size for any baseline characteristic is greater than 0.25 in absolute value, the WWC concludes that the intervention and comparison groups are not equivalent at baseline.

Figure 3: WWC standard for baseline equivalence

Absolute value of effect size ≤ 0.05	**0.05 < absolute value of effect size ≤ 0.25**	**Absolute value of effect size > 0.25**
Satisfies baseline equivalence standard	Statistical adjustment required to satisfy baseline equivalence	Does not satisfy baseline equivalence

Baseline equivalence can only be established on observable characteristics—that is, only on characteristics that can be measured. While establishing baseline equivalence reduces the threat of observable baseline differences influencing the findings, there may still be differences in unobservable characteristics that could influence the findings. That is why randomized controlled trials with high levels of attrition or other issues and quasi-experimental designs that meet the baseline equivalence standard can, at best, receive a research rating of *Meets WWC Standards With Reservations*.

If baseline equivalence is not established, the study receives a research rating of *Does Not Meet WWC Standards*.

Glossary

- The **absolute value** of a number is the size of the number without regard to its sign. For example, the absolute value of -0.10 is 0.10.
- The **analytic sample** is the sample on which the analysis is based.
- The **baseline** is the point of time prior to or early in an intervention's implementation, and ideally, at the time of assignment.
- The **effect size** is a standardized measure of the magnitude of a difference.
- The **follow-up** is the point in time during which the effect of the intervention is evaluated.
- A **standard deviation** indicates the variability of a measure across the observations in a sample.
- **Statistical adjustment** refers to including baseline measures in a statistical model to reduce bias in the estimate of the effect of the intervention.

For more information about how the WWC computes effect sizes, please see Appendix E in the *WWC Procedures and Standards Handbook*.

[2] Pooled standard deviation can be conceptualized as an "average" standard deviation, calculated across the intervention and comparison groups. For more information, please see Section F of the *WWC Procedures and Standards Handbook*.

[3] A number of different techniques can satisfy the statistical adjustment requirement, among them are regression adjustment and ANCOVA (see Chapter 3, Step 4 in the *WWC Procedures and Standards Handbook*.