

MODULE

7

Reporting

whatworks.ed.gov

MODULE

7

Reporting

This module discusses WWC reporting using the [Reading Recovery](#)® intervention report from the Beginning Reading topic area as an example.

By the end of this module, you will be able to:

- ❖ Explain how the WWC reports on the effectiveness of an intervention
- ❖ Understand the effect size and how the WWC calculates it
- ❖ Understand the improvement index and how the WWC calculates it
- ❖ Explain the statistical adjustments the WWC makes
- ❖ Explain how the WWC characterizes findings based on the effect size and statistical significance of effects for a domain

WWC Reporting

Most WWC products report on the magnitude, direction, and statistical significance of findings, but differ in how they summarize evidence from a study, and whether they provide any synthesis of evidence across multiple studies.

- ❖ **Intervention reports** synthesize all evidence about the magnitude of an intervention's effects, provide a rating of effectiveness of the intervention, and describe the extent of evidence.
- ❖ **Study findings** webpages at whatworks.ed.gov present detail about the magnitude, direction, and statistical significance of findings in an individual study.
- ❖ **Practice guides** do not present effects quantitatively in the main text; however, they present study-level findings in the technical appendix.

Key Elements for Reporting

Magnitude of Findings: Mean Difference

- ❖ The mean difference measures the impact of the intervention in the same units as the outcome measure.
- ❖ Calculated as the difference between the average outcome for intervention group members and the average outcome for comparison group members.
- ❖ May be adjusted for baseline differences by the WWC or by the study authors.

Magnitude of Findings: Effect Sizes

- ❖ An effect size represents the impact of the intervention in standard deviation units.
- ❖ Effect sizes place impacts on the same scale for different outcome measures.
 - Improves comparability across findings.
 - Allows the WWC to aggregate impacts across outcomes and studies to summarize evidence.
- ❖ The WWC may calculate an effect size twice in the course of a review to:
 1. Assess the magnitude of baseline differences (see the **Baseline Equivalence** module) and
 2. Calculate the effects of the intervention, which is the focus of this module.

Reminder: The Two Types of Effect Sizes

The WWC reports two different effect sizes depending on the nature of the outcome:

- ❖ For **continuous measures**, the WWC calculates and reports Hedges' g .
 - Calculated as the mean difference, divided by the pooled standard deviation of the outcome measure.

- ❖ For **dichotomous measures**, the WWC calculates and reports the Cox Index.
 - Yields effect sizes comparable to Hedges' g .

Calculating Hedges' g Without the Unadjusted Means and Standard Deviations

- ❖ Many study authors do not report the unadjusted means and standard deviations for their outcome measures, but instead report an impact estimate from their analysis, often adjusted for baseline measures.
- ❖ The WWC will try to calculate Hedges' g from an adjusted analysis even when unadjusted means and standard deviations are also reported because the adjustment may account for differences between the intervention and comparison groups at baseline.
- ❖ The **Study Review Guide (SRG)** has formulas to calculate the effect size from a number of common analyses. These options include:
 - Ordinary least squares (OLS), or hierarchical linear model (HLM) regression analysis; uses the unstandardized regression coefficient on the indicator for the intervention
 - Analysis of covariance (ANCOVA); can use either the adjusted post-intervention means and standard deviations, or the F -statistic for the test of the intervention effect with the pre-post correlation
 - One-way (or one-factor) analysis of variance (ANOVA); uses the F -statistic
 - Comparison of group means; uses the t -statistic

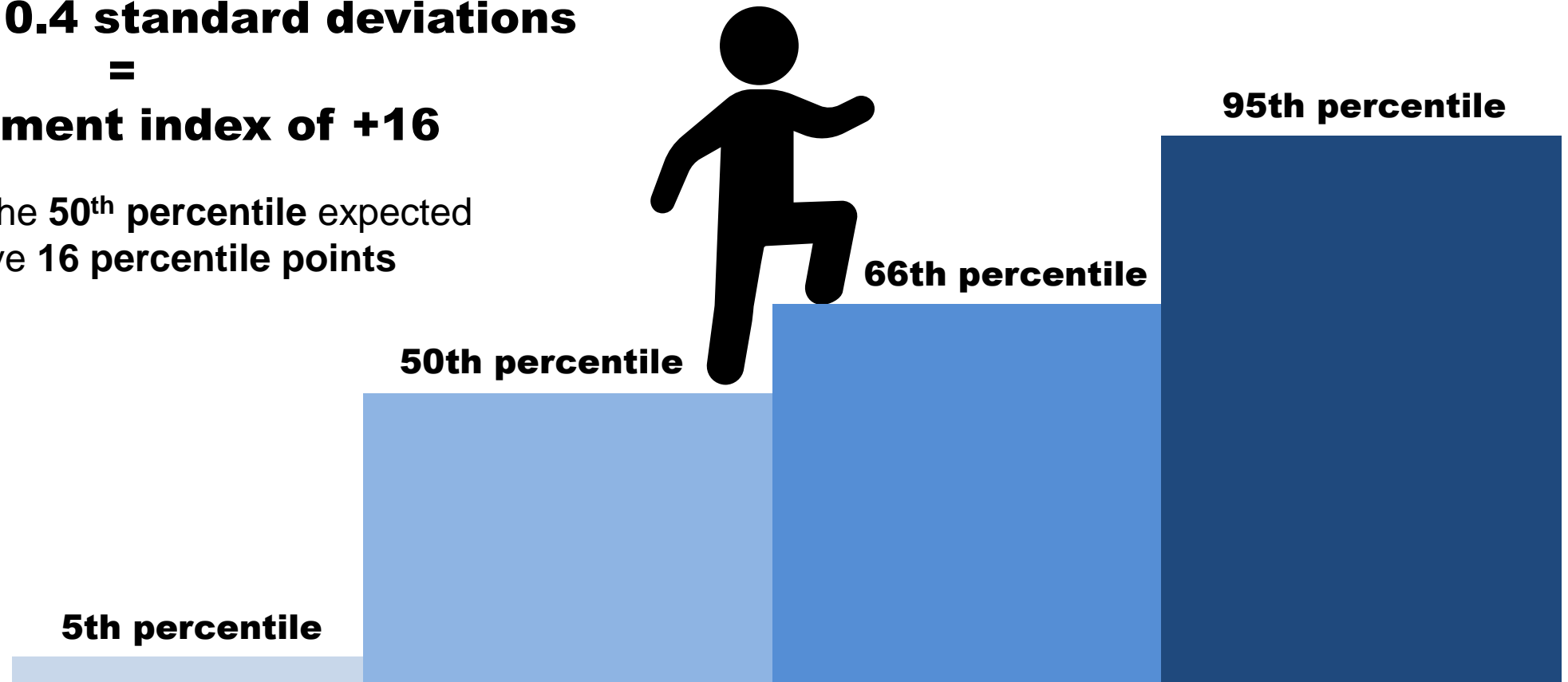
Magnitude of Findings: The Improvement Index

- ❖ The improvement index is a transformation of the effect size into percentile points.
- ❖ It reflects the expected change in percentile rank for an average comparison group student if the student had received the intervention.
- ❖ It also reflects the effect of the intervention, in percentile points, that would be expected if the comparison group's average performance was at the 50th percentile.

Example: The Improvement Index

**Effect size of 0.4 standard deviations
=
Improvement index of +16**

A student at the 50th percentile expected
to improve 16 percentile points



Statistical Significance of Findings

- ❖ An effect is described as statistically significant when the measured difference in outcomes is unlikely to have been observed under the assumption that the intervention had no impact.
- ❖ The WWC labels a finding statistically significant if the probability of observing an effect that is at least as large as the measured effect under the assumption that the intervention had no impact is less than five percent ($p < 0.05$).
- ❖ The WWC generally accepts study authors' calculations of statistical significance, with some important exceptions.
- ❖ When an acceptable calculation from the study is not available, the WWC applies an adjustment or performs its own calculation using a formula in the *WWC Procedures Handbook*.

Findings from the Reading Recovery Intervention Report

Appendix C.1: Findings included in the rating for the alphabetic domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	p-value
Construct: Phonemic awareness								
Schwartz, 2005 ^a								
Deletion task	Grade 1	74 students	6.64 (2.56)	5.58 (2.50)	1.06	0.41	+16	> 0.05
Yopp-Singer Test of Phoneme Segmentation	Grade 1	74 students	17.70 (4.93)	15.27 (5.43)	2.43	0.46	+18	> 0.05
Construct: Letter knowledge								
Pinnell, DeFord, & Lyons, 1988 ^b								
Observation Survey of Early Literacy Achievement: Letter Identification subtest	Grade 1	74 students	52.27 (1.41)	51.19 (3.17)	1.08	0.44	+17	0.06

Reporting on Findings Within Studies

MODULE

7

Reporting

- ❖ The WWC reports on findings from studies that meet the **eligibility requirements** specified in the review protocol and that meet WWC design standards with or without reservations.
- ❖ Often studies will report multiple findings, sometimes for the same outcome measure on different samples, or for multiple outcomes within a single outcome domain.
- ❖ The WWC has procedures for reporting on findings in these studies.
- ❖ Additionally, the WWC summarizes the design, features, and samples of each study to provide important context for the findings.

Reporting from Studies that Include Multiple Findings for the Same Outcome Measure

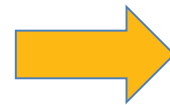
- ❖ For each outcome measure, and among those eligible findings that meet WWC design standards, the WWC designates findings as main or supplemental.
- ❖ The WWC often reports findings not designated as main findings in supplemental tables, and these supplemental findings do not contribute to the WWC summary of the evidence.

Outcomes measured at different points in time



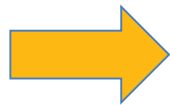
WWC review protocols identify one time point as the main finding that contributes to an evidence summary.

Full-sample or subgroups



The WWC considers the full-sample to be the main finding that contributes to an evidence summary.

Composite outcome or subscales



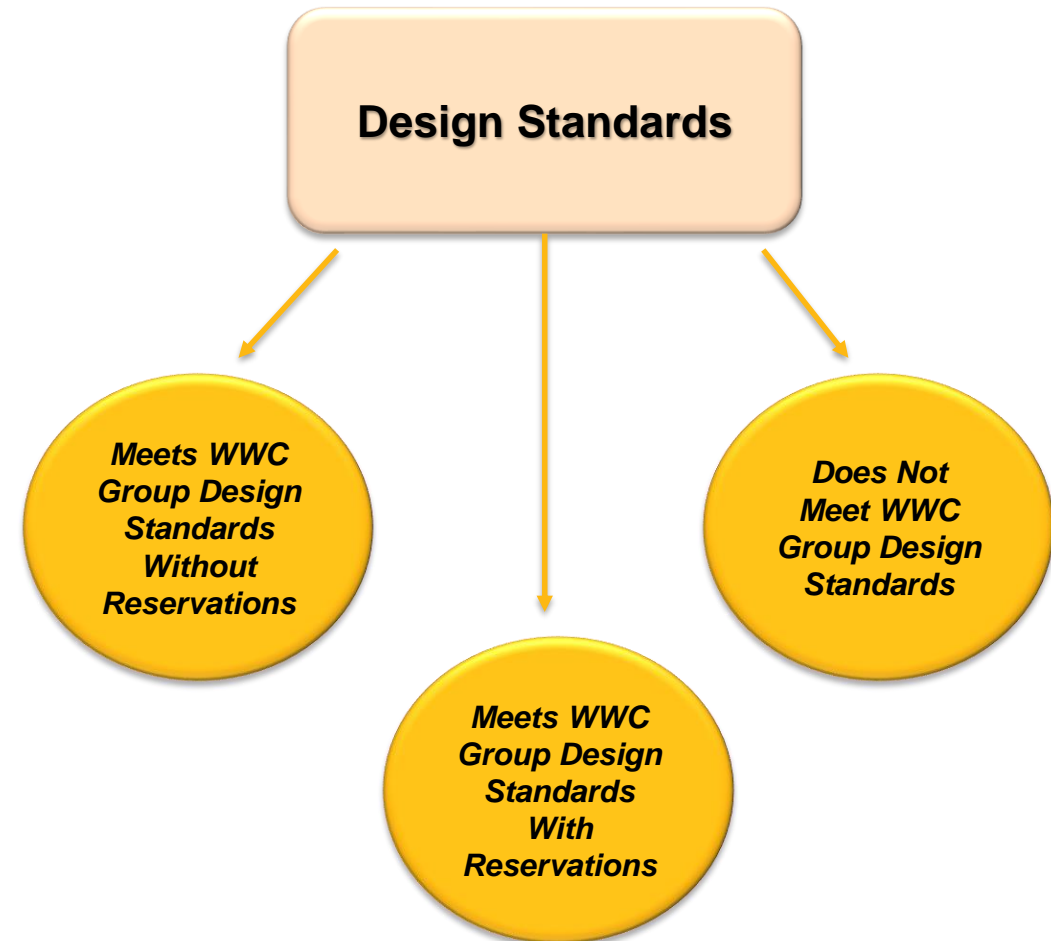
The WWC considers the composite score to be the main finding that contributes to an evidence summary.

Reporting from Studies that Include Multiple Findings for the Same Outcome Measure

- ❖ Main findings are selected among those that are eligible and *Meet WWC Group Design Standards With or Without Reservations*.
- ❖ Not all studies report a full-sample finding for the composite outcome measure and preferred time period, so review teams have discretion to identify main and supplemental findings.
- ❖ Some WWC review efforts are designated as *expedited* by IES. For expedited reviews, the review team leadership also has discretion to review eligible findings only from the full sample (rather than on subgroups), only using composite measures (rather than subscales), and only using the most relevant time period.

Rating Studies

- ❖ The WWC uses design standards to review each eligible contrast in the study.
- ❖ Different contrasts may receive different ratings.
- ❖ The study receives the highest rating among all the ratings given to eligible contrasts (main and supplementary).



Characterization of Findings at the Study Level

- ❖ The WWC characterizes main study findings within each outcome domain.

- ❖ Each outcome domain in a study is assigned one of three possible characterization labels, based on the effect sizes and statistical significance of the main findings:
 - Statistically significant positive effects
 - Indeterminate effects
 - Statistically significant negative effects

Characterization of Findings at the Study Level: One or More Outcomes?

- ❖ If there is a single outcome measure in the domain, then finding from that outcome measure determines the characterization.

- ❖ If there are multiple outcome measures in a domain, the WWC will examine each of the following:
 - The individual effect sizes and the number of statistically significant findings,
 - The simple average of the effect sizes, and a WWC-calculated measure of the statistical significance of the combined finding, or
 - A study-reported omnibus test, which is a single test that combines the outcome measures.

Positive Findings for a Domain with One Outcome Measure

- ❖ One characterization of domain-level findings indicates that the study found evidence that the intervention benefited the intervention group.

WWC Characterization of Findings of an Effect Based on a *Single Outcome Measure*

Statistically significant positive effect

The estimated effect is positive and statistically significant.

Positive Findings for a Domain With Multiple Outcome Measures

- ❖ For studies with multiple outcome measures in a domain, the domain receives a characterization of “statistically significant positive effect” if it meets one or more of three criteria.

WWC Characterization of Findings of an Effect Based on *Multiple Outcome Measures*

Statistically significant positive effect

When any of the following is true:

1. At least one main finding is positive and statistically significant, and none are negative and statistically significant based on univariate statistical tests, accounting for multiple comparisons, and correcting for clustering when not properly aligned.
2. The WWC-aggregate main finding is positive and statistically significant, correcting for clustering when not properly aligned.
3. The study reports that the omnibus effect for all outcome measures together is positive and statistically significant on the basis of a multivariate statistical test in a properly aligned analysis.

Other Characterizations of Findings

- ❖ The WWC will apply the same criteria to an outcome domain used to identify a “statistically significant positive effect” to identify those having a “statistically significant negative effect,” except that positive effect sizes are replaced by negative effect sizes.

Other Characterizations of Findings

- ❖ The WWC uses similar criteria to assign an outcome domain a characterization of “statistically significant negative effect.”
- ❖ An outcome domain receives a rating of “**indeterminate effect**” if none of the criteria for other characterizations apply.

Characterization of Study-Level Findings in the Reading Recovery Intervention Report

Below are the bottom three rows from Appendix C.1 from the Beginning Reading intervention report on Reading Recovery®. The alphabets findings in Schwartz (2005) were characterized as statistically significant positive effects.

Domain average for alphabets (Pinnell, DeFord, & Lyons, 1988)	0.47	+18	Not statistically significant
Domain average for alphabets (Schwartz, 2005)	0.62	+23	Statistically significant
Domain average for alphabets across all studies	0.55	+21	na

WWC Reviewers Document Some Information Not Affecting the Study Rating

- ❖ Documented information that does not affect the study rating:
 - Study setting
 - Characteristics of the study sample
 - Features of the intervention and comparison groups, including any implementation issues
 - How the intervention and comparison groups were formed.

- ❖ Documented information that does not affect the study rating, but may affect how to interpret a study finding and whether the study is considered a test of the intervention of interest:
 - **Non-participation.** Occurs when sample members assigned to the intervention group do not participate in the intervention.
 - **Contamination.** Occurs when sample members assigned to the comparison condition receive the intervention.
 - **Lack of Fidelity.** Occurs when components of the intervention are not implemented.

Adjustments For Reporting

Post-Hoc Statistical Adjustments

- ❖ For studies that *Meet WWC Group Design Standards (With or Without Reservations)*, the WWC prefers to report the study authors' statistical results. However, there are three situations in which the WWC applies post-hoc statistical adjustments.

- ❖ Two of these adjustments affect the statistical significance of the reported results:
 - **Clustering adjustment** (discussed in Module 8 on Cluster-Level Assignment)
 - **Multiple comparisons adjustment**

- ❖ One adjustment affects the magnitude of the findings:
 - **Difference-in-differences adjustment**

Why Make Multiple Comparisons Adjustments?

- ❖ Researchers understand that the more comparisons a study reports, the more likely it will incorrectly find that a result is statistically significant due to a **Type I error** (or false discovery).
- ❖ The WWC must guard against Type I errors, because the WWC uses statistical significance to characterize findings.



Adjustment for Multiple Comparisons

- ❖ When more findings are included, the adjustment is more severe.
- ❖ The WWC only applies the adjustment to main findings; supplementary findings are not adjusted for multiple comparisons.
- ❖ The WWC performs adjustments across main findings within an outcome domain, which will most often reflect multiple outcome measures.
- ❖ Multiple follow-ups and subgroups do not typically lead to a multiple comparisons adjustment because of how the WWC selects main findings, or aggregates findings across subgroups.
- ❖ The WWC accepts an authors' adjustment for multiple comparisons if it was performed across all main findings within the domain.

How to Make Multiple Comparisons Adjustments

The WWC uses the Benjamini-Hochberg method to adjust for multiple comparisons within a domain.

Step 1: Rank p -values

- Within a domain, rank p -values for all main findings where the smallest p -value has a rank equal to 1 and the largest p -value has a rank equal to the total number of main findings.

Step 2: Calculate the critical p -values

- For each p -value, compute a corresponding critical p -value by multiplying the rank number determined in Step 1 by 0.05 and dividing the result by the total number of contrasts in the domain.

Step 3: Determine which findings remain statistically significant after adjustment

- Find the largest p -value that is smaller than its corresponding critical p -value. The WWC considers that finding and all findings with lower rank numbers (smaller p -values) statistically significant.

Step 1: Rank p -Values

- ❖ **Example:** An author reports that four out of five outcomes within a domain were statistically significant, but did not apply a multiple comparisons adjustment. The table below ranks the five outcomes according to their p -values.

Study p -value	WWC p -value rank
0.002	1
0.009	2
0.032	3
0.033	4
0.055	5

Step 2: Calculate New Critical p -Values

In this example, the new critical p -values range from 0.01 to 0.05.

Study p -value	WWC p -value rank	New critical p -value
0.002	1	$0.01 = (0.05) * (1) / 5$
0.009	2	$0.02 = (0.05) * (2) / 5$
0.032	3	$0.03 = (0.05) * (3) / 5$
0.033	4	$0.04 = (0.05) * (4) / 5$
0.055	5	$0.05 = (0.05) * (5) / 5$

Step 3: Significance Using New Critical p -Value

- ❖ In this example, the WWC reviewer starts at the largest author-reported p -value (ranked #5) and moves up the table until reaching a p -value that is smaller than its corresponding new critical p -value. This occurs for the outcome ranked #4 (author p -value is 0.033 and critical p -value is 0.04).
- ❖ The WWC considers everything *above* the red line in the table statistically significant (even the results from the outcome ranked #3, which has an author-reported p -value higher than the new critical p -value).

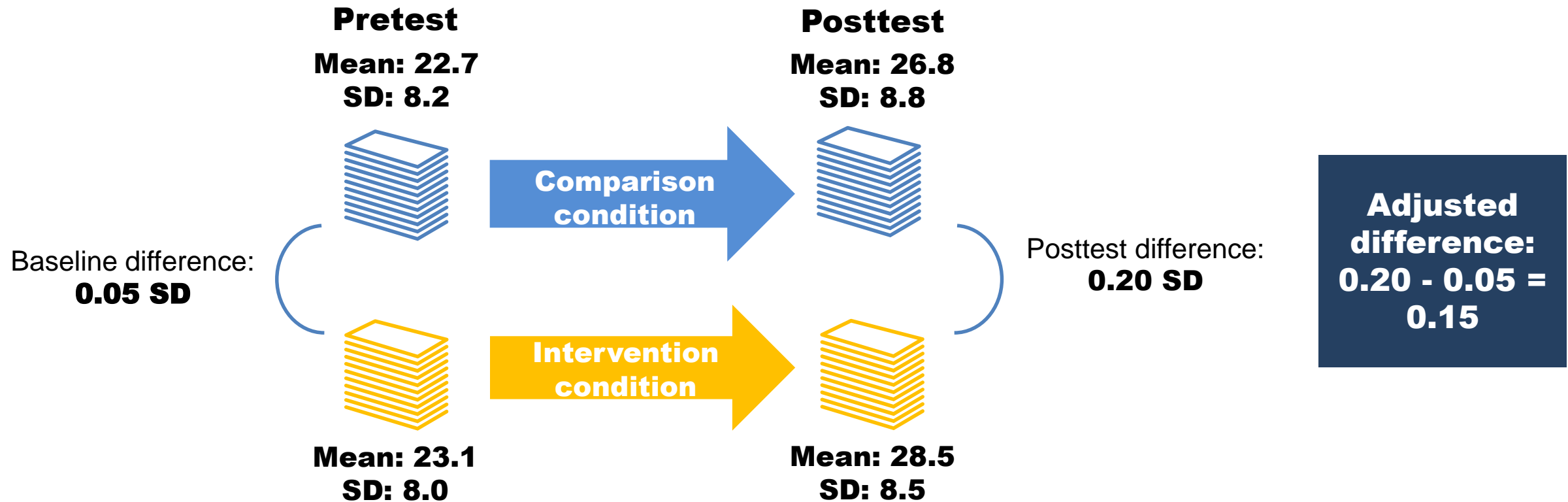
Study p -value	WWC p -value rank	New critical p -value	Is study p -value < critical p -value?	Statistically significant for WWC?
0.002	1	0.01	Yes (0.002 < 0.01)	Yes
0.009	2	0.02	Yes (0.009 < 0.02)	Yes
0.032	3	0.03	No (0.032 > 0.03)	Yes
0.033	4	0.04	Yes (0.033 < 0.04)	Yes
0.055	5	0.05	No (0.055 > 0.05)	No

Why and How Does the WWC Adjust for Baseline Differences?

- ❖ The WWC performs a difference-in-differences adjustment, to remove the pre-intervention difference between the intervention and comparison groups from the difference in outcomes.
- ❖ Even small differences can affect the magnitude of a finding.
 - **Example:** If the intervention group scored higher than the comparison group on both the pretest and the posttest, an analysis that does not correct for baseline differences may suggest an effect of the program that is too large.
 - The difference on the posttest may be due to the pre-intervention differences between conditions, rather than the intervention.
- ❖ The adjustment addresses this concern by subtracting the difference on the pre-intervention measure from the difference on the post-intervention measure.
- ❖ The WWC converts the pre-intervention and post-intervention differences to effect sizes, which allows the WWC to perform this adjustment whether the pre-intervention measure is the same or different from the outcome measure.

Why and How Does the WWC Adjust for Baseline Differences?

The WWC subtracts the pre-intervention difference from the difference in outcomes.



SD = standard deviation.

Limitation with the Adjustment for Baseline Differences

- ❖ The difference-in-differences (DND) adjustment assumes that the magnitude of the pre-intervention difference between the groups would be identical to the difference that would be measured on the post-intervention outcome in the absence of the intervention, which is unlikely.
- ❖ For this reason, the WWC reports findings from analyses in which the authors adjusted for the baseline measure when possible, rather than performing its own DND adjustment.

When Does the WWC Adjust for Baseline Differences?

- ❖ The WWC performs a DND adjustment when an analysis does not adjust for baseline differences.
- ❖ A WWC DND adjustment can be an acceptable statistical adjustment used to satisfy the baseline equivalence requirement, but only if the baseline characteristic (1) is measured using the same units as the outcome, and (2) has a correlation of 0.6 or higher with the outcome.
- ❖ When the WWC cannot perform the DND adjustment, the WWC will report unadjusted findings (except when the DND adjustment is required for the study to be rated *Meets WWC Group Design Standards With Reservations*).

Type of study that does not adjust for baseline differences	Pre-intervention effect size		
	Less than or equal to 0.05 and the authors do not adjust	Between 0.05 and 0.25 and the authors do not adjust	Greater than 0.25
RCT with low attrition	Apply WWC DND adjustment	Apply WWC DND adjustment	Apply WWC DND adjustment
RCT that must satisfy the baseline equivalence requirement	Apply WWC DND adjustment	WWC DND adjustment may allow rating of <i>Meets WWC Group Design Standards With Reservations</i>	<i>Does Not Meet WWC Group Design Standards</i>
QED	Apply WWC DND adjustment	WWC DND adjustment may allow rating of <i>Meets WWC Group Design Standards With Reservations</i>	<i>Does Not Meet WWC Group Design Standards</i>

Additional Details About the WWC DND Adjustment

- ❖ The WWC only performs the DND adjustment for a pre-intervention measure specified in the review protocol as required for baseline equivalence in the outcome domain.
- ❖ When the review protocol requires baseline equivalence on multiple pre-intervention measures, review team leadership has discretion to
 - Choose one to use for the DND adjustment, or
 - Not perform the adjustment at all if none of the available measures are thought to be sufficiently related to the outcome measure.

Reporting on Findings Across Studies

After finalizing findings from individual studies by identifying main findings and performing any necessary adjustments, the WWC summarizes findings across studies.

- ❖ **Intervention Reports** summarize findings from studies testing the same intervention
- ❖ **Practice Guides** summarize the level of evidence for recommended classroom practices or policies for educators to address challenges in their classrooms and schools

Intervention Effectiveness Rating: Across Studies

- ❖ The WWC describes the overall effectiveness of an intervention by characterizing the evidence of effectiveness across studies.
- ❖ The characterization is performed separately in each outcome domain.
- ❖ There are five possible effectiveness ratings, each based on the study-level characterizations of evidence and the study ratings:
 - **Positive effects**
 - **Potentially positive effects**
 - **Uncertain effects**
 - **Potentially negative effects**
 - **Negative effects**
- ❖ Within each outcome domain, the WWC assigns the intervention the highest effectiveness rating for which it meets the criteria.

What Are Extent of Evidence Ratings?

- ❖ The WWC reports the extent of evidence for each outcome domain, which describes how much evidence the WWC used to determine the intervention effectiveness rating.
- ❖ The extent of evidence is an indication of the external validity or generalizability of the findings.
- ❖ The WWC bases it on characteristics of the studies contributing to the intervention effectiveness rating.

Extent of Evidence Ratings

- ❖ There are two extent of evidence ratings:
 - Medium to large
 - Small

- ❖ To receive an extent of evidence rating of medium to large, a domain must meet all the following conditions:
 - Include more than one study
 - Include more than one setting
 - Have a total sample, across studies, of at least 350 students, or assuming 25 students in a class, at least 14 classrooms

- ❖ If a domain does not meet all the conditions above, it receives an extent of evidence rating of small.

The Summary Table in an Intervention Report

The front page of every intervention report includes a summary table. Below is the summary table from the Beginning Reading report on *Reading Recovery*®. The team reported on four outcome domains: alphabetics, reading fluency, comprehension, and general reading achievement.

Table 1. Summary of Findings

Outcome domain	Rating of effectiveness	Improvement Index (percentile points)		Number of studies	Number of students	Extent of evidence
		Average	Range			
Alphabetics	Potentially positive effects	+21	+9 to +42	2	148	Small
Reading fluency	Potentially positive effects	+46	+32 to +49	1	74	Small
Comprehension	Potentially positive effects	+14	+6 to +26	2	145	Small
General reading achievement	Positive effects	+27	+19 to +38	3	227	Small

Summarizing Findings Across Studies: FE Meta-Analysis

- ❖ The WWC aggregates effect sizes both within studies and overall across studies when multiple measures of the same outcome domain are available.
- ❖ Effect sizes *within studies* are aggregated using the simple average of all findings that meet WWC design standards, rounded to two decimal places.
- ❖ Effect sizes *across studies* are aggregated using a fixed-effects meta-analytic model, in which each effect size that meets WWC standards is weighted by the inverse of its variance.

Hypothetical Fixed-Effects Meta-Analysis of Intervention Impacts on Reading Comprehension

Study-Level Effects	Effect Size	SE	n	p-value
Study 1	0.92	0.24	74	<.001
Study 2	0.46	0.02	6888	<.001

Average Effect Across Studies	ES	SE	k	p-value
Meta-Analytic Average Effect Across Studies	0.47	0.02	2	<.001

Summarizing Findings Across Studies: Intervention Rating

What Works Clearinghouse

Procedures Handbook, Version 4.1

Table IV.3. What Works Clearinghouse characterization of findings in intervention reports

Characterization	Criteria
Positive effects	<ul style="list-style-type: none"> At least two studies are rated <i>Meets WWC Standards Without Reservations</i> or <i>Meets WWC Standards With Reservations</i>; AND The mean effect from a fixed-effects meta-analysis of these studies is statistically significant and positive; AND More than 50.0 percent of the fixed-effects meta-analytic weight comes from studies that are rated <i>Meets WWC Standards Without Reservations</i>.
Potentially positive effects	<ul style="list-style-type: none"> At least two studies are rated <i>Meets WWC Standards Without Reservations</i> or <i>Meets WWC Standards With Reservations</i>; AND The mean effect from a fixed-effects meta-analysis of these studies is statistically significant and positive; AND 50.0 percent or less of the fixed-effects meta-analytic weight comes from studies that are rated <i>Meets WWC Standards Without Reservations</i>. <p>OR</p> <ul style="list-style-type: none"> One study is rated <i>Meets WWC Standards Without Reservations</i> or <i>Meets WWC Standards With Reservations</i>; AND The study has a statistically significant and positive effect.
Uncertain effects	<ul style="list-style-type: none"> At least two studies are rated <i>Meets WWC Standards Without Reservations</i> or <i>Meets WWC Standards With Reservations</i>; AND The mean effect from a fixed-effects meta-analysis of these studies is not statistically

- ✓ See the Procedures Handbook for details on assigning an intervention effectiveness rating based on the fixed-effects meta-analytic average.

Knowledge Check 1

A study of an early childhood intervention measured five outcomes in the math achievement domain. The study found that four outcomes were statistically significant (operations, $p = .038$; geometry, $p = .002$; basic number concepts, $p = .021$; and number identification, $p = .032$), and one was not significant (counting, $p = .140$). The new critical values for the five outcomes are as follows: .01, .02, .03, .04, and .05.

How many outcomes remain statistically significant after adjusting for multiple comparisons using the Benjamini-Hochberg method?

- A. None
- B. One
- C. Two
- D. Three
- E. Four
- F. Five

Answer to Knowledge Check 1

- **A, B, C, D, and F are incorrect answers.** There are significant outcomes; every outcome ranked 4 or lower is statistically significant ($p = .038 < .040$). The fifth outcome is not statistically significant ($p = .140 > .05$).
- **E is the correct answer.** All four of the outcomes that the study found were statistically significant remain significant after adjusting for multiple comparisons. Rank ordering the outcomes by their p -values and their new critical p -values, we have the geometry outcome first, the basic number concepts outcome second, the number identification outcome third, the operations outcome fourth, and the counting outcome fifth. Starting with the highest p -value, the outcome for counting is not significant. Moving to the next highest p -value, the outcome for operations is significant ($.038 < .040$). Therefore, all four outcomes with a p -value rank of 4 or lower are statistically significant.

Knowledge Check 2

For each of the following situations, the authors did not include the pre-intervention measure as a covariate in their analysis. The pre-intervention measure was measured in different units than the outcome measure.

In which of the following situations would the WWC not apply a post-hoc DND adjustment?

- A. An RCT with low attrition with a pre-intervention effect size of 0.01.
- B. A QED with pre-intervention effect size of 0.04.
- C. An RCT with high attrition and a pre-intervention effect size of 0.15.
- D. An RCT with low attrition and a pre-intervention effect size of 0.29.

Answer to Knowledge Check 2

- **A is an incorrect answer.** The study is eligible to receive the *Meets WWC Group Design Standards Without Reservations* rating and shows minimal differences at baseline (0.01 standard deviations). The WWC would still apply a difference-in-differences adjustment because the authors did not control for the baseline difference in their analysis.
- **B is an incorrect answer.** The study used a QED and must demonstrate baseline equivalence. The difference is less than 0.05 standard deviations, so the study could receive the *Meets WWC Group Design Standards With Reservations* rating, even though the authors did not control for the difference in their analysis. The WWC would apply the difference-in-differences adjustment.
- **C is the correct answer.** An RCT with high attrition must demonstrate equivalence; the pre-intervention effect size is 0.15 standard deviations, so the authors must include a statistical adjustment for the pre-intervention measure in their analysis to receive the *Meets WWC Group Design Standards With Reservations* rating. The WWC's difference-in-differences adjustment cannot be used to satisfy this requirement because the pre-intervention measure was not measured in the same units as the outcome measure. Because the study did not statistically control for the pre-intervention measure, it receives the **Does Not Meet WWC Group Design Standards** rating.
- **D is an incorrect answer.** Although the pre-intervention effect size suggests that groups were not equivalent at pretest, because it is an RCT with low attrition the study does not need to use a statistical adjustment to control for the baseline difference. The WWC would apply a difference-in-differences adjustment in this situation to try to correct for the baseline difference.

Knowledge Check 3

A study reports both the full Woodcock-Johnson III Broad Reading Score and scores for the two subscales that make up the Broad Reading Score (Letter-Word Identification and Comprehension). All three findings are eligible and meet WWC design standards.

Which of the following would the WWC consider to be main finding(s)?

- A. The individual subscale scores
- B. The Broad Reading Score
- C. Both
- D. Neither

Answer to Knowledge Check 3

- **A, C, and D are incorrect answers.** The WWC prioritizes a composite score if a study reports both the composite score and the component subscale scores. In this case, the WWC would prioritize the Broad Reading score. If both a composite score and one or more component subscale scores are available, the composite score is deemed a main finding and the component scores are deemed supplemental findings.
- **B is the correct answer.** The Broad Reading Score is a composite score. The WWC would prioritize it over its component subscale scores.

Knowledge Check 4

Three studies of a middle-school math curriculum meet standards. An intervention report attempts to synthesize the following findings:

Study 1: *Meets WWC Standards Without Reservations*. It has a positive and statistically significant effect size of 0.37 and a meta-analytic weight of 0.22.

Study 2: *Meets WWC Standards Without Reservations*. It has a positive and statistically significant effect size of 0.42 and a meta-analytic weight of 0.24.

Study 3: *Meets WWC Standards With Reservations*. It has a positive statistically significant effect size of 0.54 and a meta-analytic weight of 0.54.

The meta-analytic effect size is 0.44 and it is statistically significant. How would an intervention report characterize the findings?

- A. Positive effects
- B. Potentially positive effects
- C. Uncertain effects
- D. Potentially negative effects
- E. Negative effects

Answer to Knowledge Check 4

- A, C, D, and E are incorrect answers.
- **B is the correct answer.** At least two studies are rated *Meets WWC Standards Without Reservations* or *Meets WWC Standards With Reservations*. The mean effect size from a fixed-effects meta-analysis of these studies is statistically significant and positive; and less than 50.0% of the fixed-effects meta-analytic weight comes from studies that are rated *Meets WWC Standards Without Reservations*.

MODULE

7

Reporting

Conclusion

MODULE

7

Reporting

- ❖ You can access all the resources mentioned in this module through the WWC website, **whatworks.ed.gov**.
- ❖ The full slide deck for this module, including detailed responses to the knowledge check questions, is available on the WWC website.
- ❖ To receive a certificate of completion for viewing these training modules, you must view the videos on the WWC website.

Thank you!