Version 4.1 Effect Size and Standard Error Formulas

Ryan Williams, Ph.D.
Principal Researcher
What Works Clearinghouse

David I. Miller, Ph.D.
Researcher
What Works Clearinghouse

Natalya Gnedko-Berry, Ph.D.
Senior Researcher
What Works Clearinghouse

Presenters

Ryan Williams, Ph.D.

What Works Clearinghouse

David Miller, Ph.D.

What Works Clearinghouse

Learning goals for this webinar

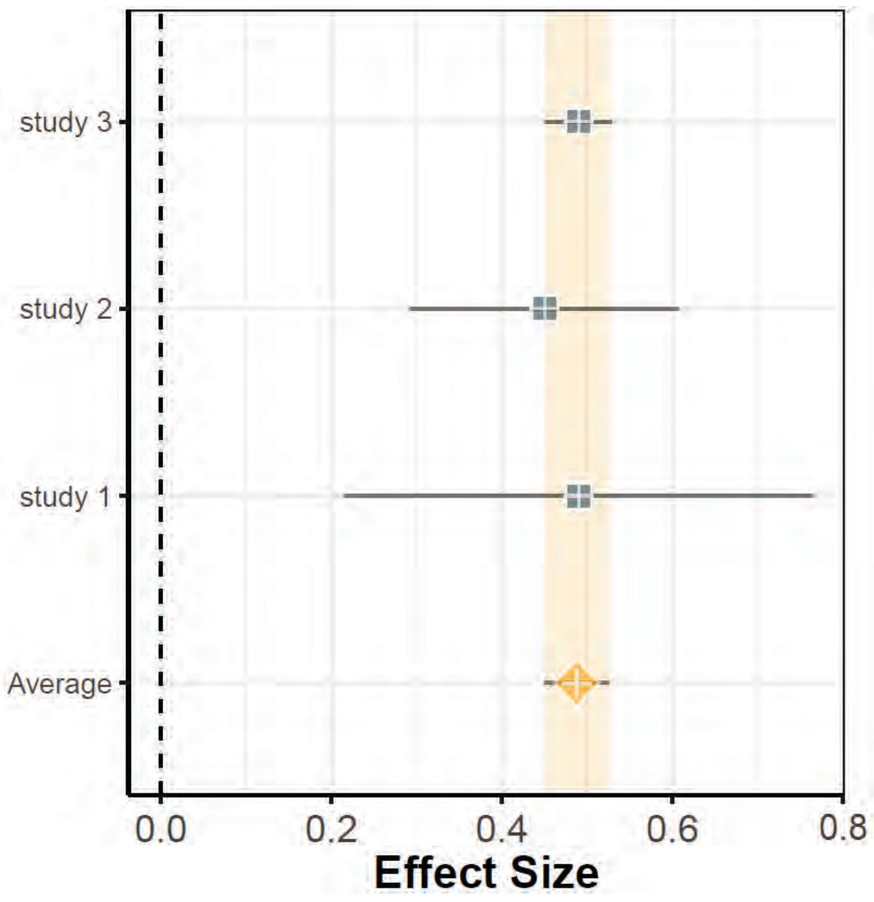
- Understand the importance of the new effect size and standard error (SE) formulas for Version 4.1 of the What Works Clearinghouse (WWC) Procedures.
- Learn the most frequently used formulas and apply them to examples.
- Understand the importance of covariate adjustment in SE calculations.
- Understand the Version 4.1 options for clustering adjustments.
- Review practical guidance for using the Online Study Review Guide to implement these formulas.

Part 1: Importance of effect sizes and standard errors in Version 4.1

1. Introduction of fixed-effects meta-analysis in Version 4.1
2. SEs to weight studies and compute *p*-values
3. Ongoing work to expand the toolbox of effect size and variance estimators

What is a fixed-effects meta-analysis?

- Fixed-effects meta-analysis involves computing a weighted average effect size.
- Studies are weighted by the inverse of the variance of their effect sizes.
- The weights are largely a function of sample size, where larger studies get proportionally more weight in the synthesis.
- A key output of this approach is a measure of precision (SE) around the mean effect, which is used, along with the direction of the average effect, to characterize interventions.



Effect sizes and standard errors in Version 4.1

- WWC transitioned to a fixed-effects meta-analytic framework in Version 4.1.
- Now WWC needs to extract information to compute effect sizes and SEs.
- The Version 4.1 supplement extends the estimation procedures to accommodate additional designs.

Effect sizes and standard errors in Version 4.1

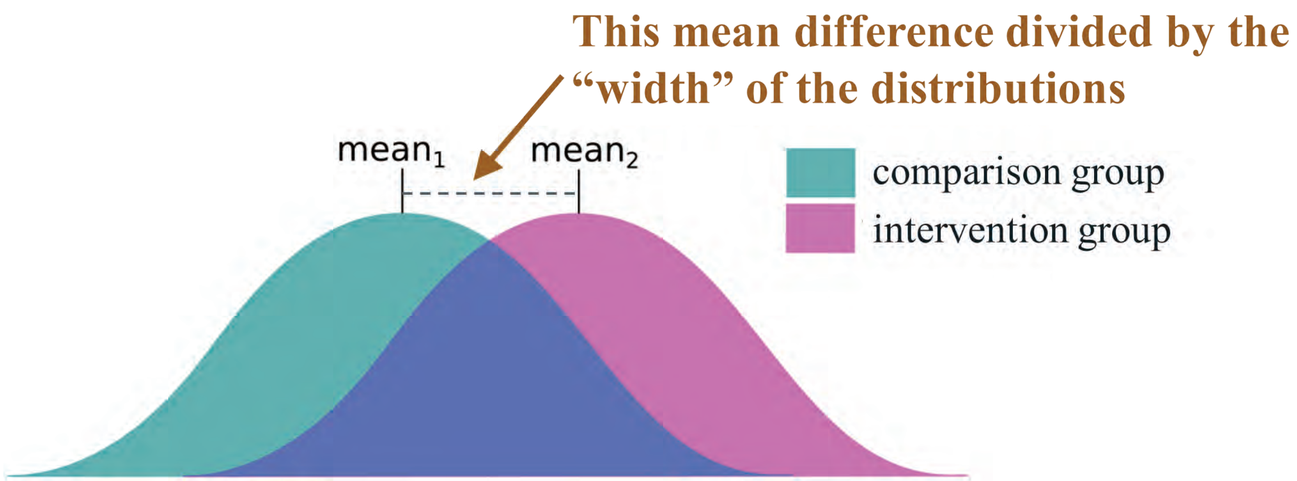
- The supplement extends the “toolbox” of effect size and SE estimators that WWC would need for different designs.
- The following sections provide an overview of the extensions that WWC developed for individual-level assignment and cluster-level assignment studies.

Part 2: Formulas for individual-level assignment studies

- Unadjusted group comparisons
- Covariate-adjusted group comparisons
- New approach for regression-adjusted SEs

Effect sizes for group mean comparisons

- WWC uses the Hedges' *g* effect size as a standardized measure of intervention effects.
  - Represents the intervention-comparison mean difference in SD units
  - Places effect sizes on a common scale for different outcome measures
  - Allows WWC to synthesize findings across measures and studies



Example information for computing effect sizes

Table F1. KeyMath3 unadjusted mean total scores

Measures	Unadjusted Mean Pretest (SD)	Unadjusted Mean Midyear (SD)
Treatment (n = 169)		
Total GSV	175.75 (15.53)	185.49 (14.72)
Control (n = 179)		
Total GSV	175.88 (15.34)	184.21 (14.67)

- Temporarily assume that only this outcome information was reported.
- Also assume that assessing baseline equivalence was not required (that is, the study is a low-attrition randomized controlled [RCT] trial).

Source: Styers, M., & Baird-Wilkerson, S. (2011). *A final report for the evaluation of Pearson's focusMATH program*. Charlottesville, VA: Magnolia Consulting.

Example information for computing effect sizes

Table F1. KeyMath3 unadjusted mean total scores

Measures	Unadjusted Mean Pretest (SD)	Unadjusted Mean Midyear (SD)
	Treatment (n = 169)	
Total GSV	175.75 (15.53)	185.49 (14.72)
	Control (n = 179)	
Total GSV	175.88 (15.34)	184.21 (14.67)

$n_i = 169$ Number of intervention students

$n_c = 179$ Number of comparison students

$y_i = 185.49$ Intervention outcome mean

$y_c = 184.21$ Comparison outcome mean

$s_i = 14.72$ Intervention outcome SD

$s_c = 14.67$ Comparison outcome SD

Computing the Hedges' *g* effect size (individual-level assignment)

1. Compute the unstandardized mean difference *b*. $b = y_i - y_c$

2. Divide by the pooled standard deviation *S*. $S = \sqrt{\frac{(n_i - 1)s_i^2 + (n_c - 1)s_c^2}{n_i + n_c - 2}}$

3. Apply the small-sample bias-correction term ω. $\omega = 1 - \frac{3}{4df - 1}$ $df = n_i + n_c - 2$

4. Generate the Hedges' *g* effect size. $g = \frac{\omega b}{S}$

Computing the Hedges' *g* effect size (individual-level assignment)

1. Compute the unstandardized mean difference *b*. $b = 185.49 - 184.21 = 1.28$

2. Divide by the pooled standard deviation *S*. $S = \sqrt{\frac{(169-1)14.72^2 + (179-1)14.67^2}{169+179-2}} = 14.69$

3. Apply the small-sample bias-correction term ω. $\omega = 1 - \frac{3}{4(169+179-2)-1} = 1.00$

4. Generate the Hedges' *g* effect size. $g = \frac{\omega b}{S} = \frac{1.00 \times 1.28}{14.69} = 0.09$

Note: For this webinar, computed values are rounded to two decimal places for presentation purposes only. To reproduce later calculations, audience members would need to use the full, unrounded values from these intermediate calculations.

Standard errors for unadjusted comparisons

- The SE represents uncertainty in the estimated effect size. Smaller values indicate more precise estimates (e.g., larger samples).
- This SE formula is applicable to unadjusted comparisons of continuous outcomes in many scenarios, including effect sizes calculated based on means or test statistics:

$$SE(g) = \omega\sqrt{\frac{n_i+n_c}{n_i n_c} + \frac{g^2}{2(n_i+n_c)}}$$

This left term is generally the bigger contributor.

This right term is usually much smaller. Represents uncertainty introduced by dividing by the sample SD.

Note: This formula is equation E.1.4 in Appendix E of the Version 4.1 WWC Procedures Handbook.

Standard errors for unadjusted comparisons: Example application

General formula:

$$SE(g) = \omega \sqrt{\frac{n_i + n_c}{n_i n_c} + \frac{g^2}{2(n_i + n_c)}}$$

Example application:

$$SE(g) = 1.00 \sqrt{\frac{169 + 179}{169 \times 179} + \frac{0.09^2}{2(169 + 179)}} = 0.11$$

Using the effect size and standard error to compute WWC *p*-values

1. Divide the effect size by the SE. $t = \frac{g}{SE} = \frac{0.09}{0.11} = 0.81$

2. Obtain the two-tailed *p*-value based on the *t*-distribution with $N - 2$ degrees of freedom.

Example Excel formula: TDIST(0.81, 169 + 179 − 2, 2)

$$p = 0.42$$

Comparing WWC-calculated and author-reported *p*-values

But WWC-calculated and author-reported *p*-values greatly differ!

	Study-reported	Calculated
Effect size	0.09	0.09
p value	0.03	0.42
Significant	Yes	No

Effect sizes match, though.

Were the authors' or WWC's calculations incorrect?

Comparing WWC-calculated and author-reported *p*-values

Were the authors' or WWC's calculations incorrect? *Neither.*

- Calculations presented so far were based on unadjusted comparisons.
- The authors' analysis controlled for a wide set of baseline covariates:
  - Baseline achievement (same measure as outcome)
  - Baseline self-efficacy
  - Demographics (e.g., socioeconomic status, race/ethnicity)
  - Teacher characteristics (e.g., years of experience)
  - School fixed effects

Why does controlling for baseline covariates matter?

Controlling for baseline covariates is favorable:

- Increases statistical precision, reducing the SE
- Can reduce bias in estimated mean differences (e.g., in quasi-experimental designs)
- Helps satisfy certain WWC baseline equivalence requirements

WWC addresses several adjustment options:

- Multiple regression (the example study used this approach)
- Analysis of covariance (ANCOVA)
- Difference-in-differences adjustment

Standard errors for covariate-adjusted comparisons

SE formula for unadjusted comparisons:

$$SE(g) = \omega \sqrt{\frac{n_i + n_c}{n_i n_c} + \frac{g^2}{2(n_i + n_c)}}$$

One acceptable SE formula for covariate-adjusted comparisons (regression and ANCOVA models):

$$SE(g) = \omega \sqrt{\frac{n_i + n_c}{n_i n_c}(1 - R^2) + \frac{g^2}{2(n_i + n_c)}}$$

More variance explained (larger R^2) translates into smaller standard errors, reflecting increased precision.

Note: The bottom formula is equation E.2.2 in Appendix E of the Version 4.1 WWC Procedures Handbook.

Standard errors for covariate-adjusted comparisons

Model R^2 values help capture the increased precision from covariate adjustment.

One problem: Studies often do not report R^2 values.

But studies often report the SE of (unstandardized) regression coefficients.

Table 10. Impact of *focusMATH* on student math achievement

Coefficient	Standard Error	*t*-Value	*p*-Value	Effect Size	Percentile Difference
1.33	0.61	2.18	0.03	0.09	4

Adjusted mean difference already used in pre-4.1 WWC versions

Critical information about uncertainty ignored in pre-4.1 WWC versions

Computing effect sizes using regression coefficients

1. Compute the unstandardized mean difference b.

$$\cancel{b = y_i - y_c = 185.49 - 184.21 = 1.28}$$

$$b = \text{unstandardized regression coefficient} = 1.33$$

Only effect size computation step that changes

2. Divide by the pooled standard deviation S.

$$S = \sqrt{\frac{(169-1)14.72^2 + (179-1)14.67^2}{169+179-2}} = 14.69$$

3. Apply the small-sample bias-correction term ω.

$$\omega = 1 - \frac{3}{4(169+179-2)-1} = 1.00$$

4. Generate the Hedges' g effect size.

$$g = \frac{\omega b}{S} = \frac{1.00 \times 1.33}{14.69} = 0.09$$

Using regression-adjusted standard errors

Table 10. Impact of *focusMATH* on student math achievement

Coefficient	Standard Error	*t*-Value	*p*-Value	Effect Size	Percentile Difference
1.33	0.61	2.18	0.03	0.09	4

How can we use this unstandardized SE to compute the standardized effect size SE?

Can we "standardize" the coefficient SE by dividing by the pooled standard deviation *S*?

- Almost but not quite.
- That would ignore the uncertainty introduced by the sample SD.
- But WWC's new regression-adjusted SE formula builds on this intuition.

WWC's new regression-adjusted SE formula

$$SE(g) = \omega \sqrt{\left(\frac{SE}{S}\right)^2 + \frac{g^2}{2(n_i + n_c)}}$$

"Standardize" the regression coefficient SE

Still needed to account for uncertainty introduced by dividing by the sample SD

Introduced by the WWC Procedures Handbook Version 4.1 Supplement in October 2020 (equation E.7.1).

- Accounts for uncertainty in both the estimated mean difference and pooled SD.
- Accounts for increased precision with covariate adjustment.
- Can even accommodate more complex regression models (e.g., multiple imputation).
- Derivation based on the delta method.

Applying the regression-adjusted SE formula

$$SE(g) = 1.00\sqrt{\left(\frac{0.61}{14.69}\right)^2 + \frac{0.09^2}{2(169+179)}} = 0.04$$

Compare the two WWC approaches to the author-reported values:

Approach	Effect size	Effect size SE	*P*-value
Unadjusted means (WWC)	0.09	0.11	0.42
Regression-adjusted SE (WWC)	0.09	0.04	0.03
Author values	0.09	–	0.03

Now the WWC-calculated and author-reported *p*-values match.

Summarizing SE formulas for individual-level assignment studies

Scenario	Effect Size SE Formula
Unadjusted comparisons	$\omega\sqrt{\frac{n_i + n_c}{n_i n_c} + \frac{g^2}{2(n_i + n_c)}}$
Covariate-adjusted model (e.g., regression, ANCOVA)	$\omega\sqrt{\frac{n_i + n_c}{n_i n_c}(1 - R^2) + \frac{g^2}{2(n_i + n_c)}}$
Regression-adjusted SE	$\omega\sqrt{\left(\frac{SE}{S}\right)^2 + \frac{g^2}{2(n_i + n_c)}}$

All the formulas shown above assume continuous outcomes. The WWC Procedures Handbook and Version 4.1 Supplement also provide formulas for dichotomous outcomes and difference-in-differences adjustments.

Part 3: Formulas for cluster-level assignment studies

- New adjustment for cluster design effect sizes
- Cluster extensions for new regression-adjusted SE approach
- Cluster extensions for other SE formulas

How WWC computes effect sizes (cluster-level assignment)

1. Compute the unstandardized mean difference *b*. $b = y_i - y_c$

2. Divide by the pooled standard deviation *S*. $S = \sqrt{\frac{(n_i - 1)s_i^2 + (n_c - 1)s_c^2}{n_i + n_c - 2}}$

3. Apply the small-sample bias-correction term ω. $\omega = 1 - \frac{3}{4h - 1}$

4. Generate the Hedges' *g* effect size, adjusting for clustered outcomes. $g = \frac{\omega b}{S}\sqrt{\gamma}$

Steps for computing the Hedges' g effect sizes for cluster-assignment studies

1. $b = y_i - y_c$

N = Total number of individuals
n = Average number of individuals per cluster

2. $S = \sqrt{\frac{(n_i-1)s_i^2+(n_c-1)s_c^2}{n_i+n_c-2}}$

Degrees of freedom

$$h = \frac{[(N-2) - 2(n-1)\rho_{ICC}]^2}{(N-2)(1-\rho_{ICC})^2 + n(N-2n)\rho_{ICC}^2 + 2(N-2n)\rho_{ICC}(1-\rho_{ICC})}$$

3. $\omega = 1 - \frac{3}{4h-1}$

Small number of clusters correction

$$\gamma = 1 - \frac{2(n-1)\rho_{ICC}}{N-2}$$

4. $g = \frac{\omega b}{S}\sqrt{\gamma}$

Standard errors for cluster-assignment studies that adjust for covariates

- Covariates in cluster-level assignment studies matter—a lot!
- The supplement adds extension to the effect size SE formulas from cluster-level assignment studies that improve precision when covariates are used and better aligns the study findings with the WWC review findings.

Standard errors for cluster-assignment studies that adjust for covariates

Here is an example of how the former WWC procedures would have produced an effect size and SE for a cluster-randomized study of an elementary school math program, by Rutherford and colleagues (2014).

Table 3. Main effect of Spatial-Temporal (ST) Math on math achievement after 1 year

		(1) Cohort 1	(2) Cohort 2	(3) Pooled	(4) 3rd Grade	(5) 4th Grade	(6) 5th Grade
ST Math	B	4.17	7.62	5.12†	3.44	6.63	5.02
	SE	(3.47)	(5.47)	(2.97)	(4.03)	(4.26)	(5.38)
	d	0.06	0.10	0.07†	0.05	0.09	0.07
Pretest Math	B	0.61***	0.60***	0.61***	0.51***	0.60***	0.74***
	SE	(0.01)	(0.02)	(0.01)	(0.03)	(0.02)	(0.03)

We would first generate our effect size estimate using the four steps outlined earlier.

Table 2. One-year differences in math California Standards Tests scores between treatment and control, divided by grade, cohort, and subgroup

		Control					Treatment				
		Pre		Post			Pre		Post		
Grade	Cohort	*M*	*SD*	*M*	*SD*	Count	*M*	*SD*	*M*	*SD*	Count
All	Pooled	349.80[a]	71.75	359.62	77.55	6,966	353.96	74.06	368.03	78.90	6,837

Source: Rutherford, T., Farkas, G., Duncan, G., Burchinal, M., Kibrick, M., Graham, J., . . . Martinez, M. E. (2014). A randomized trial of an elementary school mathematics software intervention: Spatial-temporal math. *Journal of Research on Educational Effectiveness*, *7*(4), 358–383.

Standard errors for cluster-assignment studies that adjust for covariates

1. $b = 5.12$

Table 3. Main effect of Spatial-Temporal (ST) Math on math achievement after 1 year

		(1) Cohort 1	(2) Cohort 2	(3) Pooled	(4) 3rd Grade	(5) 4th Grade	(6) 5th Grade
ST Math	B	4.17	7.62	$\boxed{5.12^{\dagger}}$	3.44	6.63	5.02
	SE	(3.47)	(5.47)	(2.97)	(4.03)	(4.26)	(5.38)
	d	0.06	0.10	$0.07^{\dagger}$	0.05	0.09	0.07
Pretest Math	B	0.61^{***}	0.60^{***}	0.61^{***}	0.51^{***}	0.60^{***}	0.74^{***}
	SE	(0.01)	(0.02)	(0.01)	(0.03)	(0.02)	(0.03)

Table 2. One-year differences in math California Standards Tests scores between treatment and control, divided by grade, cohort, and subgroup

		Control					Treatment				
		Pre		Post			Pre		Post		
Grade	Cohort	*M*	*SD*	*M*	*SD*	Count	*M*	*SD*	*M*	*SD*	Count
All	Pooled	349.80[a]	71.75	359.62	77.55	6,966	353.96	74.06	368.03	78.90	6,837

Standard errors for cluster-assignment studies that adjust for covariates

1. $b = 5.12$

2. $S = \sqrt{\frac{(6837-1)78.90^2+(6966-1)77.55^2}{6837+6966-2}} = 78.22$

Table 3. Main effect of Spatial-Temporal (ST) Math on math achievement after 1 year

		(1) Cohort 1	(2) Cohort 2	(3) Pooled	(4) 3rd Grade	(5) 4th Grade	(6) 5th Grade
ST Math	B	4.17	7.62	5.12†	3.44	6.63	5.02
	SE	(3.47)	(5.47)	(2.97)	(4.03)	(4.26)	(5.38)
	d	0.06	0.10	0.07†	0.05	0.09	0.07
Pretest Math	B	0.61***	0.60***	0.61***	0.51***	0.60***	0.74***
	SE	(0.01)	(0.02)	(0.01)	(0.03)	(0.02)	(0.03)

Table 2. One-year differences in math California Standards Tests scores between treatment and control, divided by grade, cohort, and subgroup

		Control					Treatment				
		Pre		Post			Pre		Post		
Grade	Cohort	*M*	*SD*	*M*	*SD*	Count	*M*	*SD*	*M*	*SD*	Count
All	Pooled	349.80[a]	71.75	359.62	77.55	6,966	353.96	74.06	368.03	78.90	6,837

Standard errors for cluster-assignment studies that adjust for covariates

1. $b = 5.12$

2. $S = \sqrt{\frac{(6837-1)78.90^2+(6966-1)77.55^2}{6837+6966-2}} = 78.22$

WWC default value

3. $h = \frac{[(6837+6966-2)-2(265.44-1)(20)]^2}{(6837+6966-2)(1-.20)^2+265.44(6837+6966-2\times 265.44).20^2+2(6837+6966-2\times 265.44).20(1-.20)} = 1217.93$

$$\omega = 1 - \frac{3}{4(1217.93)-1} = 1.00$$

Table 3. Main effect of Spatial-Temporal (ST) Math on math achievement after 1 year

		(1) Cohort 1	(2) Cohort 2	(3) Pooled	(4) 3rd Grade	(5) 4th Grade	(6) 5th Grade
ST Math	B	4.17	7.62	5.12†	3.44	6.63	5.02
	SE	(3.47)	(5.47)	(2.97)	(4.03)	(4.26)	(5.38)
	d	0.06	0.10	0.07†	0.05	0.09	0.07
Pretest Math	B	0.61***	0.60***	0.61***	0.51***	0.60***	0.74***
	SE	(0.01)	(0.02)	(0.01)	(0.03)	(0.02)	(0.03)

The current RCT study used random assignment at the school level. The 52 elementary schools in the study included two cohorts with a staggered implementation design.

Table 2. One-year differences in math California Standards Tests scores between treatment and control, divided by grade, cohort, and subgroup

		Control					Treatment				
		Pre		Post			Pre		Post		
Grade	Cohort	*M*	*SD*	*M*	*SD*	Count	*M*	*SD*	*M*	*SD*	Count
All	Pooled	349.80[a]	71.75	359.62	77.55	6,966	353.96	74.06	368.03	78.90	6,837

Standard errors for cluster-assignment studies that adjust for covariates

1. $b = 5.12$

2. $S = \sqrt{\frac{(6837-1)78.90^2+(6966-1)77.55^2}{6837+6966-2}} = 78.22$

3. $h = 1217.93$

$$\omega = 1 - \frac{3}{4(1217.93) - 1} = 1.00$$

4. $g = \frac{1.00(5.12)}{78.22}\sqrt{1 - \frac{2(265.44-1)0.20}{(6837+6966)-2}} = 0.07$

Table 3. Main effect of Spatial-Temporal (ST) Math on math achievement after 1 year

		(1) Cohort 1	(2) Cohort 2	(3) Pooled	(4) 3rd Grade	(5) 4th Grade	(6) 5th Grade
ST Math	B	4.17	7.62	5.12†	3.44	6.63	5.02
	SE	(3.47)	(5.47)	(2.97)	(4.03)	(4.26)	(5.38)
	d	0.06	0.10	0.07†	0.05	0.09	0.07
Pretest Math	B	0.61***	0.60***	0.61***	0.51***	0.60***	0.74***
	SE	(0.01)	(0.02)	(0.01)	(0.03)	(0.02)	(0.03)

The current RCT study used random assignment at the school level. The 52 elementary schools in the study included two cohorts with a staggered implementation design.

Table 2. One-year differences in math California Standards Tests scores between treatment and control, divided by grade, cohort, and subgroup

		Control					Treatment				
		Pre		Post			Pre		Post		
Grade	Cohort	*M*	*SD*	*M*	*SD*	Count	*M*	*SD*	*M*	*SD*	Count
All	Pooled	349.80[a]	71.75	359.62	77.55	6,966	353.96	74.06	368.03	78.90	6,837

Standard errors for cluster-level assignment studies that adjust for covariates

- Now we can estimate an effect size SE.
- We will use the unadjusted SE formula (E.5.2) first, and then we will use the formula included in the supplement (E.7.0).

Table 3. Main effect of Spatial-Temporal (ST) Math on math achievement after 1 year

		(1) Cohort 1	(2) Cohort 2	(3) Pooled	(4) 3rd Grade	(5) 4th Grade	(6) 5th Grade
ST Math	B	4.17	7.62	5.12†	3.44	6.63	5.02
	SE	(3.47)	(5.47)	(2.97)	(4.03)	(4.26)	(5.38)
	d	0.06	0.10	0.07†	0.05	0.09	0.07
Pretest Math	B	0.61***	0.60***	0.61***	0.51***	0.60***	0.74***
	SE	(0.01)	(0.02)	(0.01)	(0.03)	(0.02)	(0.03)

The current RCT study used random assignment at the school level. The 52 elementary schools in the study included two cohorts with a staggered implementation design.

Table 2. One-year differences in math California Standards Tests scores between treatment and control, divided by grade, cohort, and subgroup

		Control					Treatment				
		Pre		Post			Pre		Post		
Grade	Cohort	*M*	*SD*	*M*	*SD*	Count	*M*	*SD*	*M*	*SD*	Count
All	Pooled	349.80[a]	71.75	359.62	77.55	6,966	353.96	74.06	368.03	78.90	6,837

Standard errors for cluster-level assignment studies that adjust for covariates

- Now we can estimate an effect size SE.
- We will use the unadjusted SE formula (E.5.2) first, and then we will use the formula included in the supplement (E.7.0).

$$\text{E.5.2: } SE[g] = \omega\sqrt{\left(\frac{N_i+N_c}{N_iN_c}\right)\eta + \frac{g^2}{2h}}$$

Design effect

$$\eta = 1 + (n - 1)\rho_{ICC}$$

$$\text{E.7.0: } SE[g] = \omega\sqrt{\left(\frac{SE_{CC}}{S}\right)^2\gamma + \frac{g^2}{2h}}$$

Standard errors for cluster-level assignment studies that adjust for covariates

We already have all the information needed to compute the SE using the formula for unadjusted SEs (E.5.2).

$$\eta = 1 + (265.44 - 1)0.20 = 53.89$$

$$SE[g] = 1.00\sqrt{\left(\frac{6837 + 6966}{6837 \times 6966}\right)53.89 + \frac{0.07^2}{2\,(1217.93)}} = 0.12$$

Standard errors for cluster-level assignment studies that adjust for covariates

- To recalculate using the SE formula that is adjusted for covariates, we need to use the reported SE from the authors' outcome model.

$$SE[g] = 1.00\sqrt{\left(\frac{2.97}{78.22}\right)^2 0.99 + \frac{0.07^2}{2(1217.93)}} = 0.04$$

Table 3. Main effect of Spatial-Temporal (ST) Math on math achievement after 1 year

		(1) Cohort 1	(2) Cohort 2	(3) Pooled	(4) 3rd Grade	(5) 4th Grade	(6) 5th Grade
ST Math	B	4.17	7.62	5.12†	3.44	6.63	5.02
	SE	(3.47)	(5.47)	(2.97)	(4.03)	(4.26)	(5.38)
	d	0.06	0.10	0.07†	0.05	0.09	0.07
Pretest Math	B	0.61***	0.60***	0.61***	0.51***	0.60***	0.74***
	SE	(0.01)	(0.02)	(0.01)	(0.03)	(0.02)	(0.03)

Standard errors for cluster-level assignment studies that adjust for covariates

- The SE that is adjusted for covariates is much more consistent with the primary study results.
- The SEs that adjust for covariates result in meta-analytic weights that increase the study's contribution **by a factor of 9**.

Standard errors for cluster-level assignment studies that adjust for covariates

- While the example focused only on a case in which clustering was addressed in the primary analysis, the supplement also includes a similar approach for "mismatched analyses."

Table 2. Effect size standard error formulas based on regression model coefficients

Eq. number	Did the original study cluster correct?	Individual assignment	Cluster assignment
E.7.0	Yes	Not applicable	$\omega\sqrt{\left(\frac{SE_{CC}}{S}\right)^2\gamma+\frac{g^2}{2h}}$
E.7.1	No	$\omega\sqrt{\left(\frac{SE}{S}\right)^2+\frac{g^2}{2N}}$	$\omega\sqrt{\left(\frac{SE_{UC}}{S}\right)^2\eta+\frac{g^2}{2h}}$

Note. $\eta = 1+(n-1)\rho_{ICC}$; $\gamma = 1-\frac{2(n-1)\rho_{ICC}}{N-2}$; $h = \frac{[(N-2)-2(n-1)\rho_{ICC}]^2}{(N-2)(1-\rho_{ICC})^2+n(N-2n)\rho_{ICC}^2+2(N-2n)\rho(1-\rho_{ICC})}$; ρ_{ICC} is the intraclass correlation coefficient; S is the pooled standard deviation; SE_{CC} and SE_{UC} are the regression coefficient standard errors corrected and uncorrected for clustering, respectively; $N = n_i + n_c$, or the total sample of individuals; n is the average number of individuals per cluster.

Source: What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2020). *Supplement to the WWC version 4.1 procedures handbook.* Retrieved November 19, 2020, from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-41-Supplement-508_09212020.pdf

Other cluster-level assignment study extensions

Table 2. Effect size standard error formulas based on regression model coefficients

Eq. number	Calculation type	Individual assignment	Cluster assignment
E.2.2	ANCOVA adjusted means	$\omega\sqrt{\frac{n_i+n_c}{n_i n_c}(1-R^2)+\frac{g^2}{2(n_i+n_c)}}$	$\omega\sqrt{\frac{n_i+n_c}{n_i n_c}(1-R^2)\eta+\frac{g^2}{2h}}$
E.3.1	Gain score DnD	$\omega\sqrt{\left(\frac{n_i+n_c}{n_i n_c}\right)2(1-\rho_{cor})+\frac{g^2}{2(n_i+n_c)}}$	$\omega\sqrt{\left(\frac{n_i+n_c}{n_i n_c}\right)2(1-\rho_{cor})\eta+\frac{g^2}{2h}}$
E.3.3	Effect size DnD (no estimate of P_{cor} given—impute as 0.5)	$\omega\sqrt{\left(\frac{n_i+n_c}{n_i n_c}\right)2(1-\rho_{cor})+\frac{g_{post}{}^2+g_{pre}{}^2-2g_{pre}g_{post}\rho_{cor}{}^2}{2(n_i+n_c)}}$	$\omega\sqrt{\left(\frac{n_i+n_c}{n_i n_c}\right)2(1-\rho_{cor})\eta+\frac{g_{post}{}^2+g_{pre}{}^2-2g_{pre}g_{post}\rho_{cor}{}^2}{2h}}$
E.7.2	Effect size DnD (with estimate of P_{cor} given)	$\omega\sqrt{\left(\frac{n_i+n_c}{n_i n_c}\right)(1-\rho_{cor}{}^2)+\frac{g_{post}{}^2+g_{pre}{}^2\rho_{cor}{}^2-2g_{pre}g_{post}\rho_{cor}{}^3}{2(n_i+n_c)}}$	$\omega\sqrt{\left(\frac{n_i+n_c}{n_i n_c}\right)(1-\rho_{cor}{}^2)\eta+\frac{g_{post}{}^2+g_{pre}{}^2\rho_{cor}{}^2-2g_{pre}g_{post}\rho_{cor}{}^3}{2h}}$
E.4.4	Dichotomous outcomes	$\frac{1}{1.65}\sqrt{\frac{1}{p_i n_i}+\frac{1}{(1-p_i)n_i}+\frac{1}{p_c n_c}+\frac{1}{(1-p_c)n_c}}$	$\frac{\sqrt{\eta}}{1.65}\sqrt{\frac{1}{p_i n_i}+\frac{1}{(1-p_i)n_i}+\frac{1}{p_c n_c}+\frac{1}{(1-p_c)n_c}}$

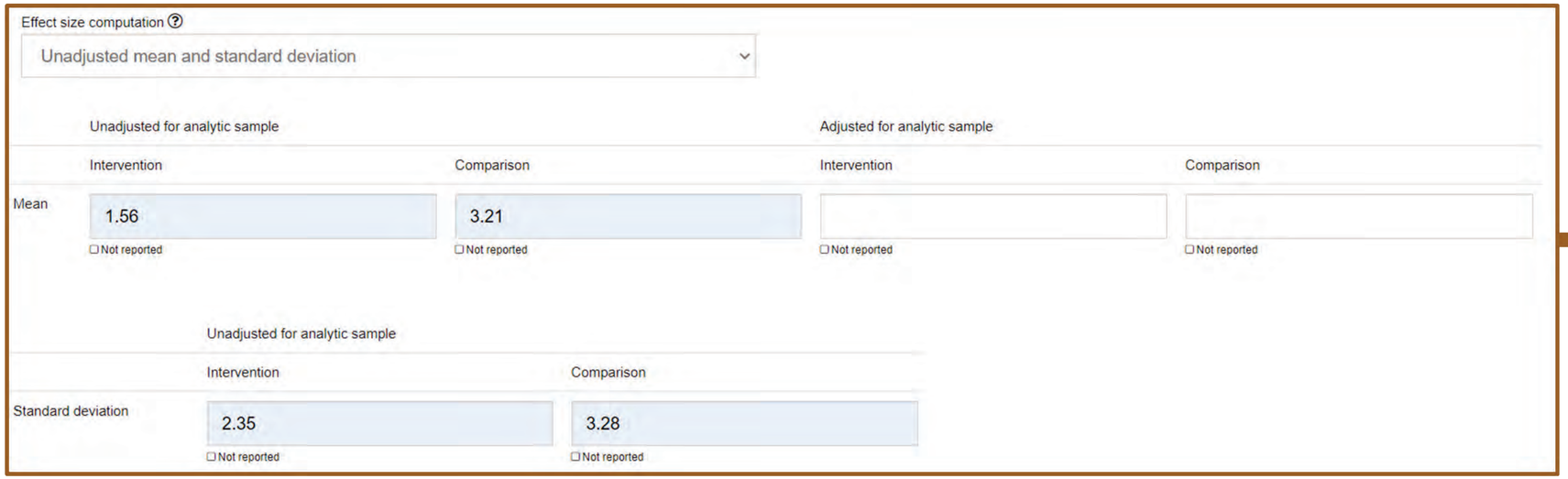
Note. $\eta = 1+(n-1)\rho_{ICC}$; $h=\frac{[(N-2)-2(n-1)\rho_{ICC}]^2}{(N-2)(1-\rho_{ICC})^2+n(N-2n)\rho_{ICC}^2+2(N-2n)\rho_{ICC}(1-\rho_{ICC})}$; ρ_{ICC} is the intraclass correlation coefficient; ρ_{cor} is the baseline-outcome correlation; $N = n_i + n_c$, or the total sample of individuals; n is the average number of individuals per cluster; p_i and p_c are the probabilities of an outcome occurring in the intervention and comparison groups, respectively. DnD = difference-in-differences.

Part 4: Practical guidance for study reviewers

- Choosing between effect size computation options
- Changes to entering statistical information in the Version 4.1 OSRG
- Understanding how entered information affects computations

Using the Online Study Review Guide for computations

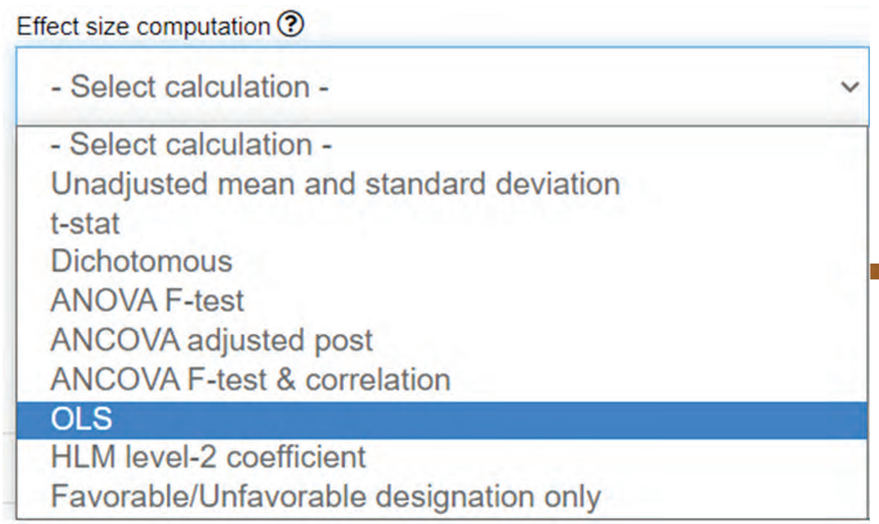
- The OSRG will automatically apply the formulas discussed today based on study information that reviewers enter (e.g., means, SDs).
- Reviewers do not need to manually calculate the effect sizes and SEs.



Generates WWC effect sizes and SEs.

Implications for WWC study reviewers

- Most Version 4.1 changes to statistical formulas affect only back-end calculations.
- But reviewers should still be aware of the expanded set of options, especially for accounting for the increased precision gains when controlling for covariates.



This selection determines which pieces of statistical information the OSRG will use in its computations.

Three principles for choosing the effect size computation option

1. **Prefer adjusted over unadjusted analyses:** Compared to unadjusted differences, analyses that adjust for baseline differences generally should be preferred, including:
   - ANCOVA adjusted post-intervention
   - ANCOVA *F*-test and correlation
   - OLS (ordinary least squares) coefficient
   - HLM (hierarchical linear modeling) level-2 coefficient

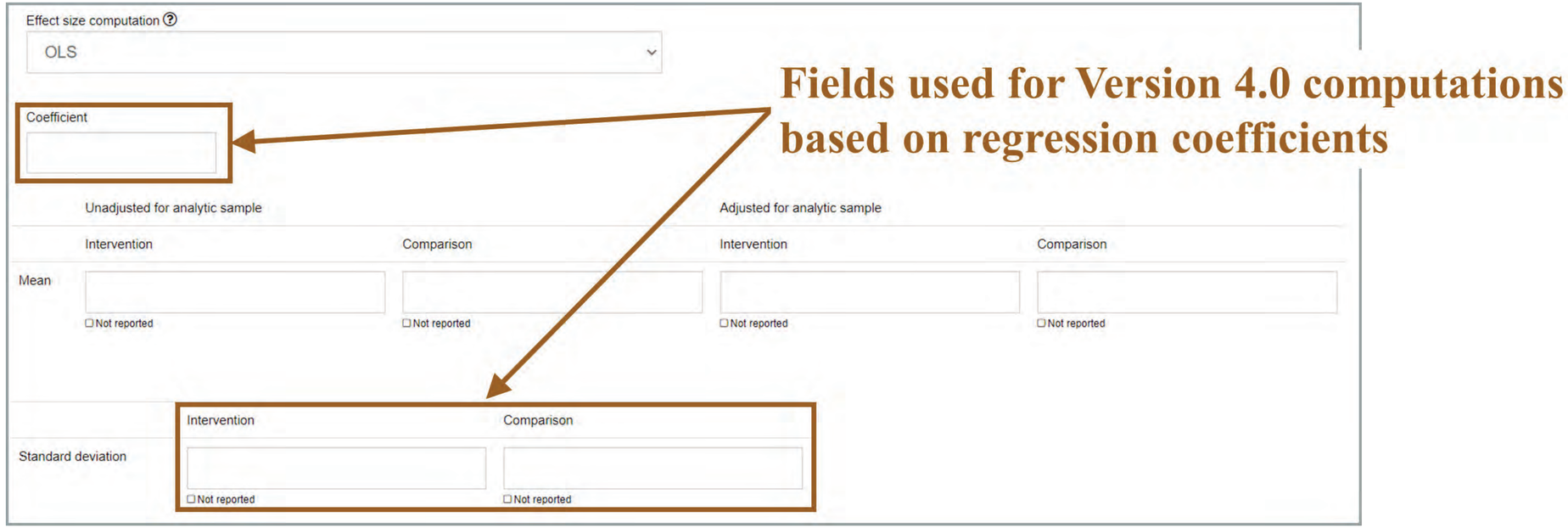
2. **The lack of reported information is often constraining:** For instance, unadjusted SDs must be entered to compute effect sizes using regression coefficients.
   - Whenever possible, it is important to select an option that allows for computing effect sizes; otherwise, the finding cannot contribute to the meta-analytic averages in cross-study syntheses.

3. **Always select "dichotomous" for dichotomous outcomes:** Dichotomous outcomes require special formulas that do not apply to continuous outcomes.

Example 1: Regression coefficient computations (old Version 4.0)

In the old Version 4.0 OSRG, reviewers could use regression coefficients (along with the unadjusted SDs) to compute WWC effect sizes and *p*-values.

But it was not possible to enter any information about uncertainty in those coefficients.



Example 1: Regression coefficient computations (new Version 4.1)

The new Version 4.1 OSRG adds three fields to implement the new SE formulas.

Old Version 4.0

Effect size computation

OLS

Regression coefficient (unstandardized)

New Version 4.1

Effect size computation

OLS

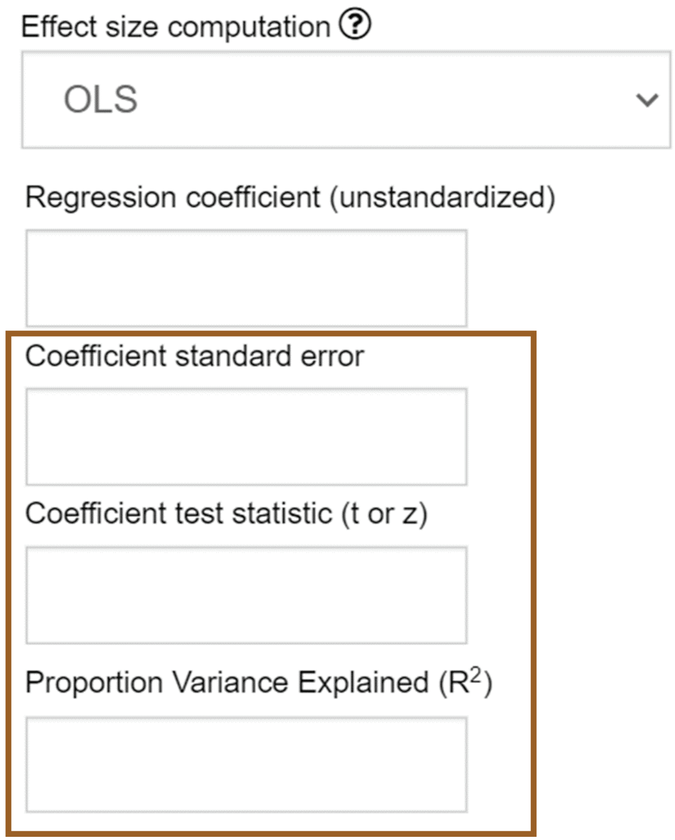
Regression coefficient (unstandardized)

Coefficient standard error

Coefficient test statistic (t or z)

Proportion Variance Explained (R^2)

Example 1: Regression coefficient computations (new Version 4.1)



These new inputs provide improved ways of calculating the SEs (and *p*-values) for WWC effect sizes.

- ✓ Accounts for precision gains with covariate adjustment
- ✓ Often makes the WWC-calculated *p*-values more closely aligned with author-reported values

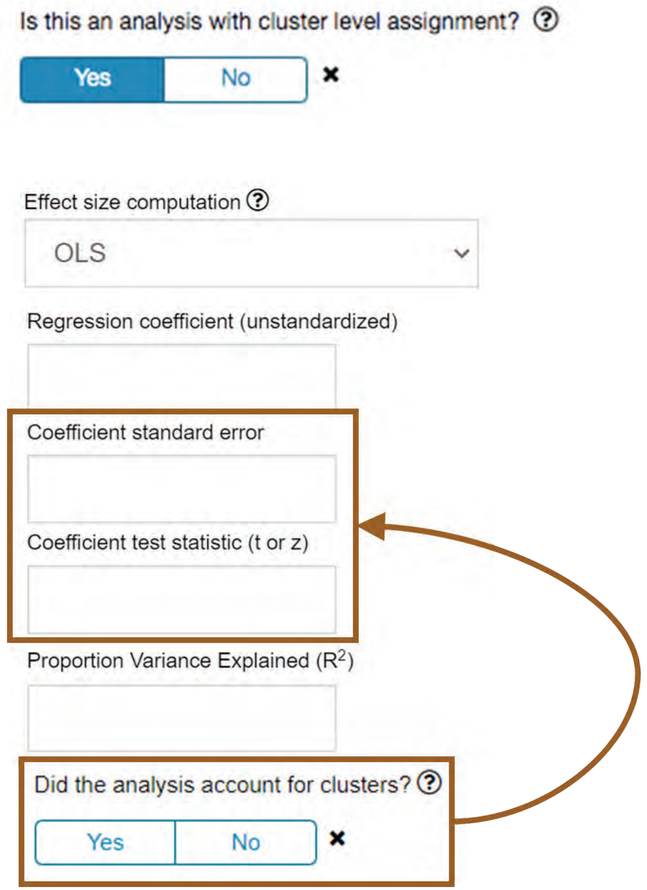
But they are not required either.

- ✓ The OSRG will automatically apply WWC defaults.
- ✓ These defaults generally are conservative ($R^2 = 0$, assuming no improvement in precision).

Example 2: Cluster corrections

WWC reviewers will flag cluster-level assignment studies.

- For some computation types, reviewers also must tell the OSRG how to interpret certain author-reported values.
  - ✓ Test statistics (*t* or *F* values)
  - ✓ Coefficient SEs



Needed to interpret other input fields correctly

Example 2: Cluster corrections with regression coefficients

Coefficient SEs generally will be much smaller if the authors' analysis did not account for clustering. The OSRG will adjust accordingly based on the reviewers' answers.

Did the analysis account for clusters? ?

Yes | No ✖

$$\omega\sqrt{\left(\frac{SE_{CC}}{S}\right)^2 \gamma + \frac{g^2}{2h}}$$

$$\gamma = 1 - \frac{2(n-1)\rho_{ICC}}{N-2}$$

This γ factor is usually very close to 1.

Did the analysis account for clusters? ?

Yes | No ✖

$$\omega\sqrt{\left(\frac{SE_{UC}}{S}\right)^2 \eta + \frac{g^2}{2h}}$$

$$\eta = 1 + (n-1)\rho_{ICC}$$

This η factor is often much larger than 1
(adjusting for regression coefficient SEs that are too small).

Practical tips for WWC reviewers

1. Take advantage of options for entering covariate-adjusted analyses, including the precision gains they offer when computing SEs and *p*-values.
2. Be aware of how the chosen effect size computation type changes what pieces of information the OSRG will use in its back-end calculations.
3. Record whether author-reported test statistics and coefficient SEs accounted for clustering in cluster-assignment studies.

A webinar slated for **January 2021** will cover these points and other changes to the Version 4.1 OSRG in more depth.

Questions?

Have questions? Contact us: https://ies.ed.gov/ncee/wwc/help

Ryan Williams, Ph.D.
What Works Clearinghouse

David Miller, Ph.D.
What Works Clearinghouse

Natalya Gnedko-Berry, Ph.D.
What Works Clearinghouse

References and further reading

- Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, *32*(4), 341–370. https://eric.ed.gov/?id=EJ782531
- Rutherford, T., Farkas, G., Duncan, G., Burchinal, M., Kibrick, M., Graham, J., . . . Martinez, M. E. (2014). A randomized trial of an elementary school mathematics software intervention: Spatial-temporal math. *Journal of Research on Educational Effectiveness*, *7*(4), 358–383. https://eric.ed.gov/?id=EJ1041346
- Styers, M., & Baird-Wilkerson, S. (2011). *A final report for the evaluation of Pearson's focusMATH program*. Charlottesville, VA: Magnolia Consulting.
- What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2020). *What Works Clearinghouse procedures handbook, version 4.1*. Retrieved November 19, 2020, from https://ies.ed.gov/ncee/wwc/Handbooks
- What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2020). *Supplement to the What Works Clearinghouse procedures handbook, version 4.1*. Retrieved November 19, 2020, from https://ies.ed.gov/ncee/wwc/Handbooks