

Questions and Answers About Effect Size and Standard Error Formulas

The purpose of this document is to provide answers to questions submitted before and during the What Works Clearinghouse (WWC) technical assistance webinar, “Updates to the Version 4.1 Online Study Review Guide.” The webinar was hosted on January 13, 2021.

This document is meant to serve as a companion to the webinar slide deck and webinar recording, both of which can be found **here on the WWC website**. We combined similar questions and rephrased others for clarity while preserving the meaning of the original questions. If additional questions arise, please contact the WWC Help Desk at <https://ies.ed.gov/ncee/wwc/help>.

1. What is the process for copying reviews conducted under Version 4.0 standards to Version 4.1 standards in the OSRG? Do these reviews need to be finalized before they can be copied?

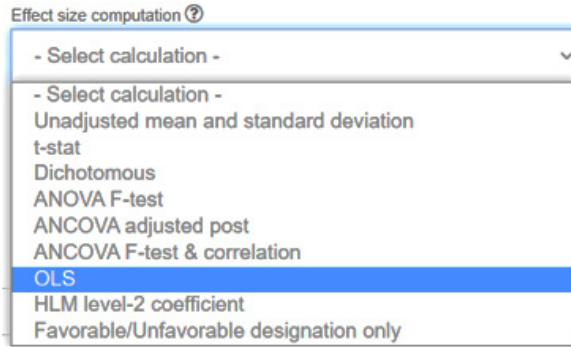
The Online Study Review Guide (OSRG) Resources Library includes a document, *How to Copy Data From Existing Reviews*, that details the review copy process. As described there, the existing review must first be finalized. Hence, if you have an in-progress 4.0 review, you must first finish it under 4.0 standards before being able to copy it for 4.1 standards. All of the 4.0 user entries will be retained, but there will be new 4.1-specific fields (such as the manual selection of attrition boundaries) for you to enter as well. Otherwise, there are no special considerations for copying from 4.0 to 4.1.

2. Can I manually select which studies will be part of a synthesis? Will the synthesis function work for practice guides that do not focus on a single intervention?

As of January 2021, there is no way to manually select individual studies or individual findings from studies to include as part of the synthesis generation process. One workaround is for reviewers to copy each study they want to include as part of the synthesis to a single intervention and outcome domain, and then run the synthesis generation process. For instance, reviewers could create an intervention name that is the practice guide name and then assign that name to copied reviews. The WWC will explore and consider enhancements to this functionality in 2021.

3. Will the Version 4.1 OSRG include notes on the Rating page describing which effect size standard error calculation was used? This would help reconcilers confirm that reviewers coded the outcome measures correctly.

In the OSRG, reviewers will choose an option in the *Effect size computation* dropdown menu that determines which pieces of statistical information the OSRG will use to compute WWC effect sizes and standard errors (see screenshot below).



This choice will be reflected in an endnote in the outcome details tables in the OSRG. The below screenshot shows an example for the ordinary least squares (OLS) regression coefficient option.

	Official	Study-reported	Calculated
Effect size	0.318 *		0.318 *
Standard error	0.139 *		0.139
Improvement index	+12		
p value	0.023		0.023
Significant	Yes		Yes

* The effect size was calculated using the OLS calculation (coefficient=7.89).

In many cases, the effect size computation option also directly corresponds to the standard error formula that the OSRG will use. In some cases, however, the OSRG’s application of standard error formulas may depend on other entered information as well.

Consider the OLS option as an example:

- If either the coefficient standard error or coefficient test statistic is entered, then the OSRG will use the regression-adjusted standard error approach noted in the supplement to the *WWC Procedures Handbook, Version 4.1* (see equations E.7.0 and E.7.1).
- If both the coefficient standard error and coefficient test statistic fields are left blank, then the OSRG will use the effect size standard error formula based on the proportion of variance explained (R^2); see equation E.2.2 in the *WWC Procedures Handbook, Version 4.1*. (The OSRG will assume $R^2 = 0$ if the R^2 field is also left blank.)
- If both the regression coefficient standard error (or coefficient test statistic) and R^2 value are entered, the OSRG will use the regression-adjusted standard error approach.

The current OSRG functionality adds a table note about the effect size computation option used (such as the OLS option), but as of December 2020, this table note functionality does not note these finer-grained standard error distinctions. Reviewers may consult appendix E of the *WWC Procedures Handbook, Version 4.1*, and its supplement to better understand the underlying logic the OSRG uses to compute standard errors.

4. If the authors provide a regression coefficient and test statistic, and do not provide an R^2 value or a standard error but do provide the means and standard deviations, what effect size option should I select and what information should I provide?

In this example, WWC reviewers should prioritize the regression coefficient, and therefore the OLS option, over the unadjusted means and standard deviation option. This is true even if the R^2 is missing; in the case of the R^2 being missing, the OSRG simply assumes a value of 0. In addition, should a test statistic *or* the standard error be available, the OSRG can use either piece of information to estimate the effect size's standard error.

As we noted in the webinar, however, reviewers should continue to fill in the means and standard deviation fields. The standard deviation fields are used to estimate the effect size. And the means may be used for reporting purposes on the WWC website or other WWC products.

5. How would you approach a situation in which the authors of a study reported an effect size, but it could not be replicated using the information provided for the study (for example, if the WWC calculated an effect size based on the information provided for the study, but it differed from the author-reported effect size)?

The WWC generally reports the WWC-calculated effect size because its computation can be verified, and it supports comparability across outcomes and studies. However, in some cases, the WWC will use the study-reported effect size if all of the following conditions are true:

- a. The study calculated the effect size consistent with a WWC formula for Hedges' g , including a small sample adjustment for continuous outcomes.
- b. The study-reported effect size was based on covariate-adjusted estimates, and the WWC-calculated effect size was not, or was based on a post-hoc difference-in-differences adjustment.
- c. The study used the Cox index for dichotomous outcomes, if appropriate.

Whether the WWC uses the study-reported effect size or the WWC-calculated effect size, reviewers should still enter the study-reported values into the OSRG. The WWC will report this supplemental information in the "Data From Individual Studies" database:

<https://ies.ed.gov/ncee/wwc/StudyFindings>.

6. Are there resources available to describe the updates and the various intricacies of the new OSRG functionality?

The WWC has created multiple resources to help WWC reviewers and review teams. The OSRG's Resource Library is available for official WWC reviewers with security clearance at <https://members.nces.ed.gov/OSRG/>; it features several documents, including the following:

- An infographic summarizing the resources for conducting Version 4.1 reviews and where to find those resources.
- A summary-of-changes document that details the new Version 4.1 features.

- An aggregating-findings document that details how to use the WWC’s new aggregation of subsample or multiple main findings procedures.

The WWC will also release a comprehensive *Version 4.1 OSRG User Guide*, targeted to both new and current users, by the end of March 2021.

The OSRG Resources Library also currently includes the *Version 4.1 Study Review Protocol*, which will later be posted on the public WWC website as well.

7. Will the WWC use the new methods to override study-reported p values?

The principles for selecting WWC-calculated versus study-reported p -values in Version 4.1 of the WWC standards remain the same as in Version 4.0. These principles state that, in contrast to study-reported effect sizes, the WWC generally accepts the study-reported p values and statistical significance (see page 9 of the *WWC Procedures Handbook, Version 4.1*).

However, there are two common circumstances in which the WWC will favor the WWC-calculated p -value over the study-reported one:

- a. The study does not include statistical significance estimates or there is a known problem with the study calculations; or
- b. The statistical significance levels reported in the study do not account for clustering when there is a mismatch between the unit of assignment and unit of analysis.

8. For the new R^2 field, should the adjusted R^2 values not be entered?

If a study reports both the raw R^2 and adjusted R^2 values, then enter only the adjusted value into the *Proportion of variance explained* OSRG input field. Adjusted R^2 values lead to more appropriate effect size standard errors by accounting for overconfidence in variance-explained estimates for models with multiple covariates. Enter 0 for negative adjusted R^2 values. However, reviewers may enter the raw R^2 value if the adjusted value is not available.

9. How often are correlation matrices reported in effectiveness studies? How should reviewers handle cases when these correlations are not provided?

Study articles will often not report correlations between outcome measures. In these cases, WWC reviewers should leave the main finding correlation input fields blank; the OSRG will automatically assume a correlation of 1 for blank entries as a conservative assumption. When study articles report those correlations, entering them can lead to improved estimation of domain-level standards errors.

10. Regarding the correlation between two outcomes, should we use the correlation between composite scores of tests (such as Woodcock-Johnson and TerraNova) or their specific subtests (such as Woodcock-Johnson reading comprehension and TerraNova Vocabulary) that were actually used in the WWC review?

The entered correlations between outcomes should correspond to the specific versions of the tests that were entered as main findings. For instance, if two subscales were designated as main findings, then the entered correlation should correspond to those two subscales.

If a study reported findings for both the composite scale and subscales, then the *WWC Procedures Handbook, Version 4.1* (pages 21–22) dictates that the composite scale should usually be considered a main finding and the subscales should be supplemental findings, potentially resulting in only a single main finding in the outcome domain. No outcome correlations would be entered if there was only one main finding (regardless of the number of supplemental findings).

However, the *WWC Procedures Handbook* also discusses instances in which the subscales should be considered main findings (for instance, if the composite scale findings were not available or did not meet WWC standards). Therefore, before deciding on which correlations to enter, it is best to consult the *WWC Procedures Handbook*.

RELATED RESOURCES

In addition to the webinar and this Questions and Answers document, the following resources provide guidance about the OSRG, effect size and standard error formulas, and the WWC’s standards, procedures, and review protocols.

- OSRG Resources Library: <https://members.nces.ed.gov/OSRG/>
- Supplement to the WWC Procedures Handbook, Version 4.1:
https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-41-Supplement-508_09212020.pdf
- WWC Standards Handbook, Version 4.1:
<https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf>
- WWC Procedures Handbook, Version 4.1:
<https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Procedures-Handbook-v4-1-508.pdf>
- WWC Data From Individual Studies database:
<https://ies.ed.gov/ncee/wwc/StudyFindings>
- WWC Help Desk:
<https://ies.ed.gov/ncee/wwc/Help>