

# Responses to comments from the public on updated version 4.1 of the *WWC Procedures Handbook* and *WWC Standards Handbook*

The What Works Clearinghouse (WWC) recently updated the [procedures and standards](#) that guide efforts to identify research on education interventions, assess the quality of the research, and summarize and disseminate evidence from studies that meet WWC standards. In January 2020, the WWC released updated versions of the *WWC Procedures Handbook, Version 4.1* and the *WWC Standards Handbook, Version 4.1*.

In July 2019, the **What Works Clearinghouse (WWC)** posted on its [website](#) proposed updates to its procedures and standards for identifying research on education interventions, assessing the quality of that research, and synthesizing evidence from studies that meet WWC standards. The proposed updates were based on consultations the Institute of Education Sciences (IES) held with the WWC’s Statistical, Technical, and Analysis Team (STAT)—which includes external consultants as well as key staff from WWC contractors—and with the WWC’s Statistics, Website, and Training (SWAT) staff at the American Institutes for Research, who led development of the WWC handbooks. IES staff from the National Center for Education Evaluation and Regional Assistance (NCEE), the National Center for Special Education Research (NCSE), and the National Center for Education Research (NCER) also reviewed the proposed updates.

The WWC also sought input from the public about the handbooks. Versions of the proposed draft handbooks were posted on the WWC website in July 2019, along with an invitation for comments from the public.

A majority of the comments received from interested members of the public focused on proposed changes to the WWC’s review process and characterization of findings from single-case design (SCD) studies. The WWC responded to the issues raised via public comment below. A summary of all changes made in version 4.1 is available on the WWC [website](#).

## **Issue 1: The WWC should consider transitioning to a random effects modeling approach instead of the proposed fixed-effects modeling approach for synthesis of studies.**

In previous versions of the *WWC Procedures and Standards Handbook*, the WWC synthesized the effects of studies in intervention reports and practice guides using a vote-counting approach in which studies were weighted equally. Under the *WWC Procedures Handbook, Version 4.1*, the WWC synthesizes the effects of studies using a fixed-effects meta-analysis. In developing the recommendation, the WWC carefully weighed the various analytic options, including using random effects estimation. We determined that the fixed-effects modeling approach is currently the best option for the WWC, based on the available research.

Implementing a random effects model requires estimating the between-studies variance component, and this estimate is poor unless the meta-analysis includes a relatively large number of studies. The fixed-effects

approach allows the WWC to conceptually acknowledge variability in effects.<sup>1</sup> Further, fixed-effects meta-analysis offers a more robust approach in comparison with the previous vote-counting approach used to synthesize studies contributing to WWC intervention reports. To date, the largest synthesis conducted by the WWC was based on nine studies, and most WWC intervention reports are far smaller: two or three studies is typical. Therefore, we believe that for the foreseeable future, it is not practical for the WWC to adopt a random effects model for its syntheses.

## **Issue 2: The WWC should reconsider removing the phrase “substantively important” as a means of characterizing intervention effects in the handbooks.**

In previous versions of the *WWC Procedures and Standards Handbook*, an effect size of 0.25 standard deviations or above was determined to be “substantively important” within the context of the WWC. Interventions that did not have a statistically significant positive impact, but that did have an effect size meeting this threshold, were eligible to receive a “potentially positive” intervention effectiveness rating.

The decision to remove the “substantively important” designation was motivated by the WWC’s decision to revise procedures for assigning effectiveness ratings to interventions. The new system relies on a fixed-effects meta-analysis of effects from individual studies, and the assessment of the sign and statistical significance of that average effect, rather than a vote counting approach relying on individual studies suggesting positive effects. The new emphasis on pooling effect size estimates across studies increases the power of statistical hypothesis tests and does not incorporate judgments about the magnitude of the observed effects into judgments about intervention effectiveness. The U. S. Department of Education’s [evidence definitions](#) adopted in 2017 also emphasize an effect’s sign and statistical significance, rather than its magnitude, as a necessary criterion for strong, moderate, or promising evidence of effectiveness. For these reasons, the “substantively important” designation was no longer necessary or helpful for the WWC to use in its reviews and syntheses of evidence. In addition, the “substantively important” designation was applied without regard to study context. This meant that, for example, a study with an effect size of 0.25 on homework completion and a study with an effect size of 0.25 on high school graduation had the same designation, but most people would say that graduation is a far more important outcome than completing homework.

## **Issue 3: The WWC should restore visual analysis to the procedures for making an initial characterization of the findings from single-case design studies meeting WWC standards.**

In the *WWC Standards Handbook, Version 4.0*, visual analysis was applied to SCD studies that met standards to determine “whether evidence of a relation between an independent variable and an outcome variable exists” (p. A-6). If at least 5 SCD studies that met WWC standards were conducted by at least 3 non-overlapping study teams and included at least 20 cases (the “5-30-20” rule), then the WWC used those studies to form effectiveness ratings for an intervention on the basis of a “vote counting” [approach](#). However, absent a method of calculating effect sizes from SCD studies, the findings from SCD studies could not be synthesized with findings from group design studies to establish overall effectiveness ratings for interventions.

---

<sup>1</sup> We wish to clarify the difference between “fixed-effect” and “fixed-effects.” Briefly, a fixed-effect meta-analysis assumes that the studies in a single synthesis are estimating a common effect size. A fixed-effects model does not invoke this assumption; instead, it assumes that the interest is in restricting inferences to studies that are highly similar to the studies in the meta-analytic database.

In July 2019, the WWC proposed that its new *Procedures Handbook, Version 4.1* specify the use of a design-comparable effect size (D-CES) for SCD studies for which this effect size could be calculated appropriately. The percentage of SCD studies previously meeting WWC pilot SCD standards that permit the calculation of a D-CES is greater (70 percent) than the percentage included in WWC intervention reports under the “5-3-20” rule (53 percent). Relying on a D-CES allows the WWC to characterize the statistical significance of findings from SCDs. The characterization of statistical significance in SCDs is needed for two reasons: first, to inform the effectiveness ratings included in WWC intervention reports, and second, to differentiate between studies which provide strong, moderate, or promising evidence of effectiveness under the U. S. Department of Education’s [evidence definitions](#). In addition, the use of a D-CES allows the WWC to assign a single effectiveness rating and an improvement index to an intervention in each outcome domain on the basis of a fixed-effects meta-analytic framework. Therefore, the WWC can now include findings from eligible SCDs and group design studies in effectiveness ratings and an improvement index. While visual analysis will not be used to characterize study findings, WWC reviewers will collect this information if it is provided by authors for inclusion in the Study Review Guide.

**Issue 4: For reviews of SCD studies, the WWC should consider adding language from the Version 4.0 Reviewer Guidance document on how to use visual information when rating a SCD study and determining whether and how to calculate a design-comparable effect size.**

Some reviewers of SCD studies under the WWC’s pilot SCD standards noted that visual information may be helpful in reviews of SCD studies, even if the WWC does not use visual analysis to characterize intervention effects. In particular, the [Version 4.0 Reviewer Guidance for Use With the Procedures and Standards Handbook](#) outlines visual information that WWC reviewers should use when reviewing SCD studies. Based on the comments the WWC has received from SCD reviewers, much of the information in the reviewer guidance document has been incorporated into the SCD section of the *WWC Standards Handbook, Version 4.1* to help WWC reviewers rate SCD studies. The relevant content includes, for example, guidance on residual treatment effects, understanding concurrence of treatment phases in multiple baseline and multiple probe designs, rating designs with more than the minimum amount of data needed to meet standards, and the inclusion of the changing criterion design as eligible to be rated under the reversal/withdrawal standards. Given the value of visual information for WWC reviews of SCD studies as well as to SCD researchers, an SCD study for which raw study data in either graphical or tabular format is unavailable will be rated *Does Not Meet WWC SCD Standards*.

The section of the *WWC Procedures Handbook, Version 4.1* describing the calculation of the D-CES also specifies that the WWC will use visual information to determine whether and how to calculate D-CES values from SCD studies meeting WWC SCD standards. For example, the WWC’s default assumption when estimating D-CES values from studies using multiple baseline designs is to assume no trend at baseline or any later phases. Review team leadership may determine, on the basis of visual inspection or an appropriate algorithm, that the underlying data do not conform to these specifications. The review team may, after consultation with the content and methodological experts, either change the specifications or not compute the D-CES, if an appropriate method is not available. The WWC will document the rationale for any departures from the default specifications for computing the D-CES in the Study Review Guide.

**Issue 5: The WWC should consider adding several omitted designs currently used in SCD research.**

The WWC recognizes that some designs from SCD research have not been included in version 4.1 or in previous versions of the handbooks. We anticipate adding more designs and effect size estimators from those designs in either handbook supplements, which would be specified in study review protocols, or in future updates to the handbooks.

**Issue 6: The *WWC Procedures Handbook, Version 4.1*, should explicitly state modeling decisions that need to be made to estimate the design-comparable effect size.**

As a result of consultation with several SCD experts and with NCSE staff, the final *WWC Procedures Handbook, Version 4.1*, includes several additional details on how a D-CES will be estimated. This includes using restricted maximum likelihood estimation, treating intervention effects as fixed effects, and assuming there is no trend at baseline for multiple baseline designs. Future handbook supplements or handbook updates may include additional modeling details, should new or updated effect size estimation techniques be added.

**Issue 7: The *WWC Procedures Handbook, Version 4.1*, does not specify how the WWC will treat SCD studies where no effect size can be estimated appropriately.**

An SCD study must be rated *Meets WWC SCD Standards With or Without Reservations* before the WWC will estimate an effect size from that study. Should a study author only provide data in graphical format, the WWC will extract raw data from those graphs for use in effect size computation. Apart from these considerations, the WWC will treat SCD studies for which a D-CES cannot be estimated appropriately in the same way the WWC treats other study designs when an effect size cannot be estimated: the WWC will list such studies and their study ratings in relevant intervention reports or practice guides and in the WWC [reviews of individual studies database](#), but will not use those studies to determine intervention effectiveness ratings or to justify a “strong” or “moderate” base of evidence for a practice recommendation.