

Response to Comments from the Public on Updated WWC Procedures and Standards

The What Works Clearinghouse (WWC) has recently updated the procedures and standards that guide our efforts to identify existing research on education interventions, assess the quality of the research, and summarize and disseminate evidence from studies that meet design standards. In October 2017, two updated handbooks were released: *WWC Procedures Handbook (Version 4.0)* and the *WWC Standards Handbook (Version 4.0)*.

In this multi-year process of updating the WWC's standards and procedures, the Institute of Education Sciences (IES) consulted with the WWC's Statistical, Technical, and Analysis Team (STAT)—which includes outside consultants as well as key staff from different WWC contractors—to develop these Handbooks. IES also gathered input from anonymous peer reviewers and from methodology experts on topics including regression discontinuity designs, cluster-level assignment, missing data, and complier average causal effects.

The WWC also sought input from the public through posting drafts of updated standards for regression discontinuity designs in December 2015 and for cluster designs in April 2016. Versions of the proposed new handbooks were posted on the WWC website in August 2017, along with an invitation for comments from the public.

The WWC received a number of suggestions from the public, many of which were incorporated into the Version 4.0 Handbooks. A **summary of all of the changes** between Version 3.0 and Version 4.0 is available on the WWC website.

However, some of the suggestions require further consideration and deliberation. In all of its work, the WWC follows a rigorous, systematic, and transparent approach—and considering or making any changes or updates to either the standards or procedures is no exception. The effort is a continuous one, as new techniques are brought into analytical practice. As a result, although we have sets of procedures and standards that are released at a specific point in time, they are also undergoing continuous assessment for possible future updating. As in the past, any proposed update to WWC standards or procedures in the future will include the opportunity for public comment on proposed changes to the Handbooks.

Some important issues raised by commenters did *not* lead to changes in the Version 4.0 Handbooks, but will inform the future development of WWC standards, procedures, and review protocols. We describe three of these issues below.

Issue 1: Why hasn't the WWC incorporated into the Handbooks advances in research methods such as single-case designs?

Although the 4.0 standards expanded upon or incorporated some previously excluded methods, we recognize that the approaches used by researchers to measure the impacts of interventions are constantly evolving with advances in methodological research. The updates to WWC procedures and standards reflect an attempt to keep pace with those changes and provide additional clarifications for readers and users of the Handbooks. As research methodology continues to evolve, the WWC must balance the development of new standards with ensuring that the additional methods are tested and stable, and that appropriate WWC standards are developed for research designs that may be distinct from the research designs currently eligible for WWC review.

For this update, the WWC both revised existing standards and developed new standards to account for advances in research methods in the field of education. These Handbooks include substantial updates and clarifications for statistical adjustments for demonstrating equivalence, methods for dealing with missing data, and analyses of studies with cluster designs. Additionally, the regression discontinuity design standards have been expanded, and new standards were developed for complier average causal effects.

However, we recognize that methodological advances are ongoing—and that there are areas of the Handbooks that will be identified for further refinement. As an example, the WWC introduced pilot standards for reviewing single-case designs with Version 3.0 of the Handbook; however, these pilot standards were not updated in this version of the Handbooks. (The WWC has issued additional **guidance** on reviewing studies that use these designs, based on input solicited from single-case design experts. We expect to update and finalize the standards for single-case designs in a future handbook revision.) Broadly speaking, as additional techniques are developed and validated and these practices are adopted more widely in the education research community, the WWC strives to ensure that our standards and procedures adhere to current best practice.

Issue 2: Why does the WWC accept researcher-developed outcome measures?

Studies reviewed by the WWC measure the effects of interventions using a variety of outcome measures, including standardized tests, constructs from administrative data (such as student attendance or dropout rates), and researcher-developed measures. However, some researchers may be concerned that findings based on researcher-developed measures may show larger effects of interventions compared to established standardized tests. For example, a researcher-developed measure evaluating skills on a narrow set of math skills may find a larger effect than a standardized test that covers a broad set of math topics, including a subset of questions on those skills.

Procedures and standards in place under the previous version of the standards to address this concern have been retained in the new Handbooks. In addition, WWC **review protocols**, that guide all WWC reviews, specify which outcome measures are eligible for review. Additionally, the *Standards Handbook* lists several requirements that outcome measures must meet, including face validity, reliability, and lack of *overalignment* with the intervention. An outcome measure that is overaligned is one that assesses content that is tailored to the intervention condition. For example, a measure that included the same text passages that were used as part of the intervention condition, but were not seen by students in the comparison condition, could be considered overaligned.

These WWC requirements for outcome measures will exclude some researcher-developed measures when they do not measure skills that are aligned with the focus of the review protocol, have not been shown to be sufficiently reliable, or appear to be designed to show effects of the intervention through construction of the outcome measure, familiarity with the format, or other means besides learning educationally relevant material. However, other researcher-developed measures that meet the requirements for outcomes will continue to be included in WWC reviews.

Issue 3: Why does the WWC continue to use the “substantively important” designation?

The WWC characterizes a finding as substantively important if it is larger than a specified threshold, but not statistically significant. This designation was created to highlight potentially effective interventions that may have been evaluated in studies that were not large enough to find a statistically significant effect. This characterization of findings was designed in order to prevent the WWC from overlooking promising evidence and from focusing primarily on studies that show very small impacts based on large samples. While both the statistical significance and substantive importance of findings contribute to how the WWC characterizes an intervention in a WWC intervention report, statistical significance carries more weight in part because of concerns about inferences from studies based on small samples. At least two studies must have a positive and statistically significant finding for the intervention to receive the highest intervention rating of *positive effects*. But, under Version 4.0 procedures, a single study with a positive and substantively important finding can continue to qualify an intervention for the next highest intervention rating of *potentially positive effects*. The WWC is considering alternative approaches to characterizing and synthesizing study findings, which may affect the continued use of this designation in future versions of the *WWC Procedures Handbook*.