

Key Criteria Used in WWC Reviews of Single-Case Design Research

Single-case designs (SCDs) are experimental designs that generally include a small number of participants and have the potential to demonstrate causal effects. SCDs must adhere to the same requirements in terms of the general outcome requirements. However, when SCD studies are focused on causal effect estimates, they differ from group designs in how they generate causal effect estimates. As a result, SCD studies require a different review process with different specific standards from group designs. SCD studies may also contain more than one experiment, and each experiment should receive its own rating.

In 2015, the WWC worked with a panel of experts to develop criteria for determining the rating of effectiveness for an intervention, based on the single-case design studies that met WWC single-case design standards. This panel also helped develop [Reviewer Guidance for Use with the Procedures and Standards Handbook](#).

In 2022, the WWC provided updated guidance on how to rate SCDs with features from multiple design types and SCDs with more cases and/or phases than the minimum required to meet WWC standards. Under version 5.0, a study is typically eligible to receive the highest rating that any subset of cases or phases is eligible to receive. The WWC made this change to ensure that SCDs that include information above and beyond what is required by the standards are not penalized for reporting more data than studies that report the minimum data required, and to allow study authors more flexibility to design studies using a combination of design features. This change also brings SCD study ratings into closer alignment with group design study ratings.

When reviewing and reporting on single-case design research, the WWC determines:

- the rating of single-case experiments and studies that include them
- the rating of effectiveness for an intervention

Eligibility for WWC review of an SCD

The eligibility criteria for a WWC review of SCDs require that the study be publicly available, released within the 20 years preceding the review, uses eligible populations, examines eligible interventions, and has eligible outcomes. In addition, studies that are eligible for review as SCDs are identified by the following features:

1. An individual case is the unit of intervention administration and data analysis. A case is most commonly a single participant. It also may be a cluster of participants, such as a classroom or school.
2. Within the design, the case can provide its own control for purposes of comparison. For example, the case's series of repeated outcome measurements prior to the intervention is compared with the series of repeated outcome measurements during and after receiving the intervention.
3. The outcome variable is measured *repeatedly* within and across different conditions. These different conditions are frequently structured as phases, such as the first baseline phase, first intervention phase, second baseline phase, and second intervention phase.

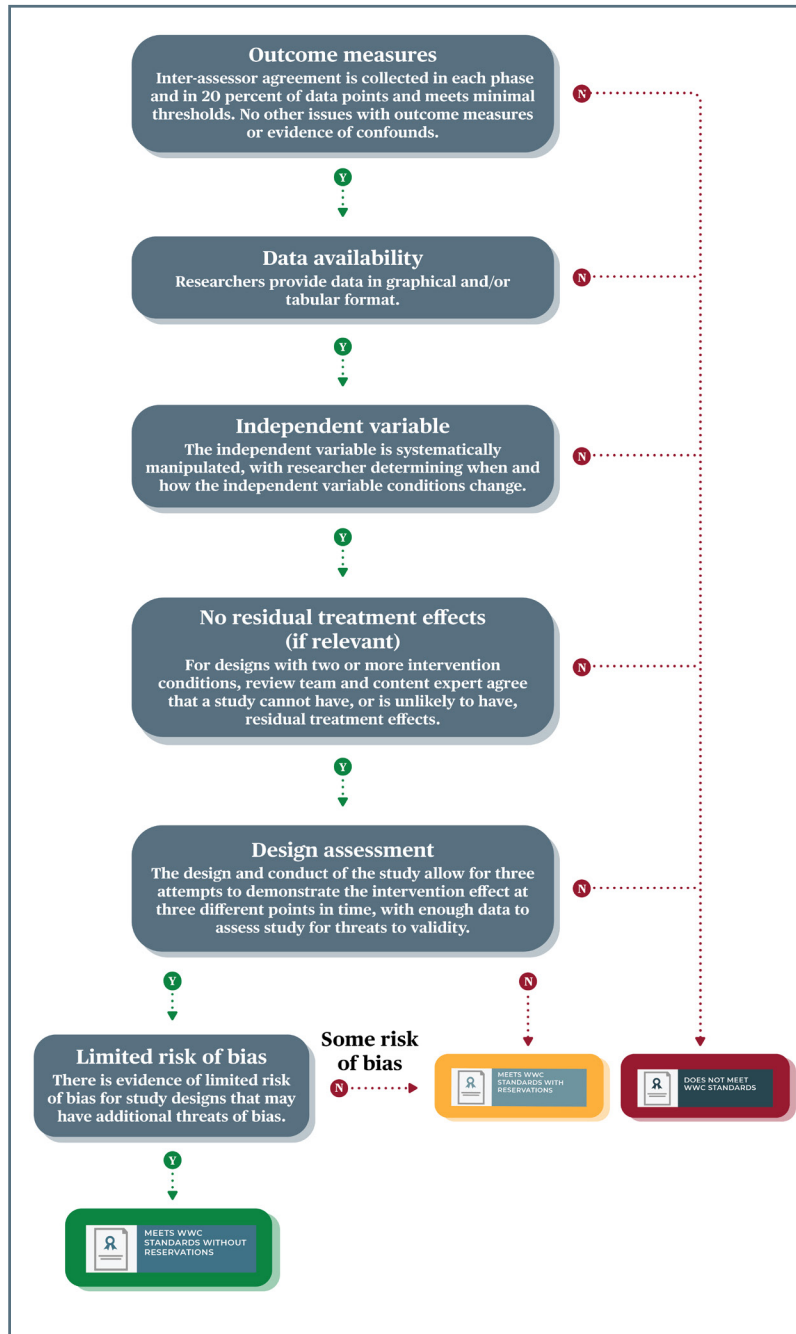
WWC standards apply to a wide range of SCDs, including reversal/withdrawal designs, multiple baseline designs, alternating and simultaneous intervention designs, changing criterion designs, and variations of these core designs like multiple probe designs.

Under version 5.0, single-case design studies (SCDs) that use multiple baseline/multiple probe, treatment reversal, and changing criterion designs need to have at **least six data points in the initial baseline phases** for their findings to be eligible for the rating *Meets WWC Standards Without Reservations*.

Reviewing findings from SCDs according to WWC standards

The process for reviewing SCD studies that are found eligible for review is presented in **figure 1**. After a study is found eligible for a WWC review as an SCD, the next step is the same as in other designs and includes reviewing the study's outcome measures and checking for confounding factors. If none of the outcome measures are consistent with the WWC's standards or if the study contains a confounding factor, the study will receive a research rating of *Does Not Meet WWC Standards* and the review will stop.

Figure 1. Single-case design review process for eligible study findings¹



1. Please note that figure 1 is actually figure 15 in the *What Works Clearinghouse Procedures and Standards Handbook*, Version 5.0 on page 106.

Outcome measure standards

The WWC's outcome measure standards for SCDs are similar to those for randomized controlled trials (RCTs) and quasi-experimental designs (QEDs) described in Chapter III, Outcome measure standards, including (1) face validity, (2) reliability, (3) not overaligned with the intervention, and (4) consistent data collection procedures. Differences in these standards for SCDs are described next.

Standard 1: Face validity

The requirements for face validity are the same as those for group designs. To show evidence of face validity, an outcome measure must appear to measure what it claims to measure. To demonstrate face validity, a measure must have a clear definition of what it measures, such as a skill, an event, a condition, or an object, and assess that skill or event. For instance, a measure described as a test of reading comprehension that only assesses reading fluency does not demonstrate face validity.

Standard 2: Reliability

For a measure to demonstrate reliability, study authors must present evidence that the outcome has acceptably low levels of measurement error. In group designs, study authors typically report measures of internal consistency, temporal stability, or test-retest reliability. In SCDs, outcomes are most frequently direct observations of behavior. For these direct observation outcomes, the most applicable form of reliability is interassessor agreement, also known as inter-rater reliability or interobserver agreement. Although more than 20 statistical measures can represent interassessor agreement (for example, see Berk, 1979; Suen & Ary, 1989), commonly used measures include the percentage or proportional agreement and Cohen's kappa coefficient, which adjusts for the expected rate of chance agreement (Hartmann et al., 2004). Minimum acceptable values of interassessor agreement are at least .80 if measured by percentage agreement, and at least .60 if measured by Cohen's kappa (Hartmann et al., 2004).

To meet the WWC's interassessor agreement requirements for direct observation outcomes, the following criteria must be met:

1. The outcome variable must be measured systematically over time by more than one assessor for each case.
2. The study authors must collect interassessor agreement in each phase.
3. The study authors must collect interassessor agreement data for at least 20 percent of the data points.
4. The interassessor agreement must meet the minimum acceptable values for each outcome across all phases and cases (however, the interassessor agreement values are not required to meet minimum acceptable values separately for each case or phase). The raw data from the secondary assessor that were gathered for the purposes of interassessor agreement do not need to be reported. It is enough to report summary measures of interassessor agreement.

If a study contains measures that are not direct observations of behavior, such as a test of an academic outcome, then the reliability standards for these measures will follow the guidelines in Chapter III, Outcome measure standards.

Standard 3: Not overaligned

Overalignment occurs when an outcome measure contains content or materials provided to the cases in one condition but not another. This rule does not apply when material covered by an outcome measure must be explicitly taught, or when an outcome measure is broadly educationally relevant. Content experts can provide advice on whether an outcome has broad educational relevance. These two caveats to the overalignment standard are particularly important to SCDs, which frequently focus on narrow, specific outcomes that may require explicit teaching, or on daily-living outcomes with educational relevance. The functional skills domain from the [Study Review Protocol](#) contains examples such as dressing, preparing and eating food, or hygiene, where the researcher might teach the participant a checklist or a set of steps that need to be repeated and the outcome might be some measure of success at repeating the checklist or steps that were taught to the participant.

Standard 4: Consistent data collection procedures

Data must be collected in the same manner for the intervention and comparison conditions. If no information is provided, the WWC assumes that data were collected consistently. In the context of SCDs, the reviewer should ensure that the data collection procedures were similar across conditions for a given case. Reviewers should look for details indicating that data were collected in different modes, with different timing, or by using different personnel in the different conditions.

In terms of timing, the major concern is whether the data collection takes place at a different time of day between conditions. For instance, if all baseline data points are collected in the morning but the intervention data points all are collected in the afternoon, this would represent inconsistent data collection procedures. However, in many SCDs, the introduction of the intervention is staggered in a time-lagged fashion across participants in the design. Staggered introduction of an intervention that is an intentional element of the design does not represent an issue with inconsistent data collection procedures.

Additional consideration: Independence of outcome measure

The consideration for independence is unchanged for SCD designs. That is, in some outcome domains as specified in the [Study Review Protocol](#), the WWC will consider whether the measure is independent of the intervention.

Criteria used to determine the effectiveness rating(s) for an intervention

When the threshold for reporting single-case design research has been reached for a given outcome domain, the WWC summarizes the body of evidence in an intervention report, using a rating of effectiveness. For each domain, this effectiveness rating for the intervention is based on all of the single-case design experiments presented in the studies that meet WWC pilot single-case design standards. The rating is based on the consistency of demonstrated effects of the intervention across all single-case design experiments.

Updated Guidance for Rating SCDs

In Version 5.0 of the Handbook, the WWC provided updated guidance on how to rate SCDs with features from multiple design types and SCDs with more cases and/or phases than the minimum required to meet WWC standards. A study is typically eligible to receive the highest rating that any subset of cases or phases is eligible to receive. The WWC made this change to ensure that SCDs that include information above and beyond what is required by the standards are not penalized for reporting more data than studies that report the minimum data required, and to allow study authors more flexibility to design studies using a combination of design features. This change also brings SCD study ratings into closer alignment with group design study ratings.