



WHAT WORKS CLEARINGHOUSE

Study Review Guide Instructions for Reviewing Single-Case Designs Studies

Released May 2016
Updated May 26, 2016

This document provides step-by-step instructions on how to complete the Study Review Guide (SRG, Version S3, V2) for single-case designs (SCDs).

Reviewers will complete an SRG for every What Works Clearinghouse (WWC) review. A completed SRG should be a reviewer's independent assessment of the study, relative to the criteria specified in the review protocol and the WWC Procedures and Standards Handbook (Handbook). At the end of the review process, a Master SRG (MRG) will represent the final assessment of the study, provide a record for the WWC, and serve as a key input in producing reports. For more details on the processes for particular reviews, see the instructions provided in the review protocol.

As central members of the review team, reviewers are expected to conduct high quality reviews. The documentation for each review should be clear and grammatically correct. Each review product should also be as complete as possible, as reviewers are accountable for the accuracy of each SRG that they prepare. Keep in mind that MRGs are deliverables and can be made publicly available.

This guide is intended to be used by people trained and certified in the WWC pilot SCD review standards and procedures and in conjunction with the WWC Handbook. See the resources available at <http://ies.ed.gov/ncee/wwc/resources.aspx> for the links to instructions, the Handbook, and the most current version of the SRG.

Definition of a Study in SCD Research

The WWC characterizes all SCD experiments in the same publication as one study, and thus one study/publication can have multiple SCD experiments. For example, a study could include three separate ABAB experiments for one student when there are data for three different eligible outcomes. Alternatively, a study can include three separate ABAB experiments for three separate eligible students when there are data for one eligible outcome. Each eligible SCD experiment should receive a WWC design rating (or disposition code) in the SRG. The study will also receive an overall study rating that the citation list will report.

Navigating and Adding Rows to the SRG

There are four worksheets (that is, tabs) in the SCD SRG: (1) Main, (2) Data, (3) Author Query & Response, and (4) Summary. Reviewers might have to add rows to the *Main Tab* to capture multiple dispositions and information about multiple SCD experiments or outcomes within a study. They might also have to add rows to the *Data Tab* and/or *Author Query & Response Tab*.

To add additional rows in any worksheet, click “Unprotect Sheet” under the *Review* options in the Microsoft Excel toolbar/ribbon. Copy the rows you want to add by selecting them with your cursor, right clicking, and selecting “Copy.” Place your cursor in the row below where you want to add them, right click, and select “Insert Copied Cells.” Always use the “Copy” and “Insert Copied Cells” functions to preserve text, drop-down menus, and formatting in the SRG. Repeat the process to add as many rows as you need in the various tabs. When you are done adding rows, click “Protect Sheet” to ensure you do not overwrite a formula by mistake.

Main Tab of the SRG

The *Main Tab* of the SRG captures a prose summary and assessment of the study, structured in four separate stages: (1) preliminary screening, (2) quality of evidence, (3) evidence rating, and (4) study details. We describe each stage in detail next.

Reviewers should provide a complete response to each question in each stage (unless a question does not apply to a specific design or a stage is not required because none of the experiments use a design that meets WWC SCD Pilot Standards). In general, reviewers will be required to enter data into columns C, D, and E in the *Main Tab*.

- Column C typically requires a short answer to a question posed in column B.
- Column D requires a justification for the short answer in column C.
- Column E requires the page numbers from the study that serve as the source of the justification presented in column D.

Stage 1: Preliminary Screening

Stage 1 of the study review assesses whether a study is eligible for WWC review under a given review protocol. All reviews are conducted under a specific review protocol. All screening and review decisions relate to that specific protocol and could differ under another protocol. For instance, a study that is ineligible under one protocol could be eligible under another because the sample or outcome measures align with the latter but not the former.

- **Study ID.** (cell C2)
Enter the Study ID for the review.
- **Reviewer Number.** (cell C3)
Enter your reviewer number. For an MRG, enter “MRG” and list all reviewer numbers (MRGs should not contain any names).
- **Review Date.** (cell C4)
Enter the date of your review.

- **Full Citation.** (cell D2)

Enter the full citation for the study.

Example:

Bettinger, E., Long, B., Oreopoulos, P., & Sanbonmatsu, L. (2012). The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment. *Quarterly Journal of Economics*, 127(3), 1205–1242.

Overview

- **Standards and Protocol.** (row 9)

Enter the version of standards used for the review (cell C9). Enter the name and version of the protocol used for the review (cell D9).

- **Intervention Name.** (row 10)

Clearly indicate the name of the intervention(s) you are reviewing within this SRG (cell C10). Enter the exact name of the intervention and a brief description from the study (cell D10). Note if the intervention has multiple versions or names.

Example:

(Cell D10) *The study examined the effectiveness of Reading Recovery implemented as a pullout program for struggling readers. Students participated in 30-minute lessons four days a week.*

Screening Criteria

The responses in this section can be brief. If the study is rated either *Meets WWC Pilot Single-Case Design Standards without Reservations* or *Meets WWC Pilot Single-Case Design Standards with Reservations* (in Stage 2), Stage 4 of the SRG will provide a more detailed description of the intervention, study participants, and setting.

- **Effectiveness.** Does the study examine the effect of an intervention? (row 13)

In column C (cell C13), select “Yes” if the study claims to examine the effect of an intervention within the scope of the review, regardless of the quality of the design; select “No” otherwise.

- **Design.** Does the study use a single-case design? SCDs have the following features: (1) an individual case serves as the unit of intervention administration and data analysis, (2) the case provides its own control for purposes of comparison, and (3) the outcome variable is measured repeatedly within and across different conditions or levels of the independent variable. (row 14)

In column C (cell C14), select “Yes” if the study uses an SCD (that is, it meets the conditions in cell B14). Select “No” if the study does not use an SCD.

In column D (cell D14), select the type of design from the drop-down menu. Design options include:

- Single Case Design (SCD)
- Randomized controlled trial (RCT)
- Cluster RCT
- Quasi-experimental design (QED)
- Regression discontinuity design (RDD)

If the design is a comparison group design or RDD, stop entering information into the SCD SRG template. If you are certified to conduct a review for the given design, obtain the appropriate SRG template for the comparison group design or RDD and complete the SRG in the correct template. If you are not certified in the applicable design, contact review team leadership.

- **Focus.** Is the intervention a program, product, policy, or practice aligned with the review protocol? (row 15)

In cell C15, select “Yes” if the intervention meets the criteria for inclusion specified in the protocol under *Types of Interventions to be Included*; select “No” otherwise.

In cell D15, briefly describe the intervention as it will be reviewed within the SRG.

- **Sample Alignment.** Does the study meet the requirements for sample characteristics specified in the review protocol? (row 16)

In cell C16, select “Yes” if the intervention meets the criteria for inclusion specified in the protocol under *Types of Populations to be Included* (for example, percentage English language learners or percentage general education); select “No” otherwise.

In cell D16, describe the sample with respect to how it satisfies or does not satisfy this requirement.

Examples:

(Yes) Each child had a diagnosis of autistic disorder, provided by an outside professional, according to the criteria of the American Psychiatric Association DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, 4th edition) (APA, 1994). Diagnoses were confirmed using the Autism Diagnostic Interview, Revised (ADI-R, Lord, Rutter, Sr Le Couteur, 1994) and the Autism Diagnostic Observation Schedule (ADCS, Lord, Rutter, DiLavore, & Risi, 1999).

(Yes) Donna was an 8-year-old girl with severe intellectual disability and cerebral palsy; Denise was a 9-year-old student with a severe intellectual disability who was legally blind. Ralph was a 6-year-old student with a severe intellectual disability and cerebral palsy.

- **Time.** Was the study published within the time frame established in the review protocol? (row 17)

In column C (cell C17), select “Yes” if the study falls within the time frame outlined in the protocol; select “No” otherwise.

- **Age or Grade Range.** Does the study examine students in the age or grade range specified in the review protocol? (row 18)

In column C (cell C18), select “Yes” if the intervention meets the criteria for the age or grade range specified in the protocol under *Types of Populations to be Included*; select “No” otherwise.

In column D (cell D18), provide (1) the grade levels of the students in the sample; and/or (2) the age range of the children in the sample, if the sample describes

them in terms of age rather than grade level. Provide both grade and age information, if available.

Example:

(Yes) *Six participants were 2 to 4 years old (their ages ranged from 24 to 47 months).*

- **Location.** Does the study examine sample members in a location specified in the review protocol? (row 19)

In column C (cell C19), select “Yes” if the study drew its sample from the geographic region described in the protocol under *Types of Populations to be Included*; select “No” otherwise. In column D (cell D19), describe the location.

- **Outcomes.** Does the study include at least one eligible outcome in a review protocol domain? (row 20)

In column C (cell C20), select “Yes” if the study examines the effects of the intervention on at least one outcome that falls into one of the domains specified in the protocol under *Types of Outcomes to be Included*; select “No” otherwise. In column D (cell D20), list all outcomes included in the study, note which are eligible for review, and cite the relevant domain. You do not have to provide detailed information for each outcome here.

Example:

(Yes) *The reading rate and reading accuracy outcomes were both in the reading fluency domain. The teacher proficiency outcome is not eligible for this review.*

Select “No” if there are no outcomes that fall within the domains of the review protocol. In column D (cell D20), briefly list the outcomes from the study. Provide any additional documentation needed for a reviewer to confirm that the outcomes are not eligible under that protocol. If you are part of a WWC team and are uncertain whether any of the study’s outcomes qualify under the protocol, seek additional guidance from review team leadership.

Example:

(No) *The outcome is not eligible for review under the Students with Learning Disabilities review protocol. Although the study examined reading fluency, the outcome was measured in Spanish.*

- **Does the study meet WWC and review protocol screening criteria?** (row 22)

In column C (cell C22), select “Yes” if the study met all of the screening criteria in this section and warrants a full review to determine if the study design meets the WWC’s standards. In cell D22, select “The study is eligible for review. An explanation for that decision is below.” from the drop-down menu.

Select “No” if the study failed to meet any of the screening criteria in this section, and select the appropriate screening disposition code in column D (cell D22).

If a study is ineligible for more than one reason, copy the row and insert additional dispositions as needed.

The screening disposition codes are:

... **is out of scope of the protocol.** This covers scenarios in which the study did not (1) use an intervention conducted in English, (2) occur within the time frame specified in the protocol, (3) occur within a geographic area specified in the protocol, (4) occur within a setting specified in the protocol, or (5) include an outcome within a domain specified in the protocol.

... **does not use a sample aligned with the protocol.** This covers scenarios in which (1) the study did not use a sample within the age or grade range specified in the protocol or (2) [the study did not show/the authors could not confirm/the WWC could not confirm] that at least 50% of the sample was classified [review specific classification].

... **does not use an eligible design.** This covers scenarios in which the study did not (1) examine the effectiveness of an intervention; (2) contain a primary analysis; (3) use a comparison group, regression continuity, or single-case design; or (4) provide adequate or consistent information to assess whether it was eligible for review.

- **Explanation for screening disposition.** (row 23)

If you selected “No” in cell C22, provide a complete explanation for the screening disposition you selected in cell D22.

Examples:

(Yes) The study is eligible for review under the Intellectual Disabilities review protocol. The study is an SCD that examines the impacts of customized checklists on general housekeeping tasks for high school-age students diagnosed with intellectual disabilities.

(No) The study is not eligible for review under the Children Identified with or at Risk for an Emotional Disturbance review protocol because the study participants were 4 years old and attended preschool at the time of the study.

Stage 2: Quality of Evidence (if the study passes Stage 1)

Stage 2 of the review determines whether the study meets WWC standards. In this section, you will assess whether the design(s) in the study meet WWC standards. This stage is organized into several sections: (1) graphical illustration(s) of the data, (2) information about the study design(s) and unit of analysis, (3) outcome(s) and outcome requirements, (4) active manipulation of the independent variable, (5) residual effects, (6) attempts to demonstrate effects, (7) additional criteria for multiple baseline and multiple probe designs, (8) additional criteria for multiple probe designs, and (9) other design issues.

Whenever possible, paraphrase the information from the study and provide details to improve transparency of the review process. Use direct quotes, with quotation marks, from the study as needed. If, at any stage of the review, information from the study is unclear, document this and indicate that an author query might be necessary. Draft all questions you have on the *Author Query & Response Tab*.

Information about graphical presentations of data

- **Does the study present a graphical illustration of the data for at least one experiment?**

In cell C28, use the drop-down menu to select “Yes” if the study presents a graphical illustration of the data (for example, using a figure) for at least one experiment. Otherwise, select “No.”

Information about the Study Design(s) and Unit of Analysis

- **Does the study use a reversal/withdrawal, multiple baseline, multiple probe, alternating treatment, changing criterion, or other design? Information about the study design and unit of analysis.** (rows 30 through 36)

In cells C30 through C35, use the drop-down menus to select “Yes” or “No” for each design: Reversal/Withdrawal (C30); Multiple Baseline (C31); Multiple Probe (C32); Alternating Treatment (C33); Changing Criterion (C34), Other (C35). If you select “No” for any design(s), the options in the corresponding drop-down lists in cells C52 through C55, C58 through C61, and C63 through C65 will be limited to “Not Applicable.” Describe any notable features about the applicable design in cells D30 through D35.

In some cases, an SCD experiment could simultaneously use more than one type of design (such as a multiple baseline design that could also be treated as three separate ABAB designs). In this case, you should consult with review team leadership to decide which design the WWC will use to characterize the findings. Select the design you have decided to use in this section, and include a note about the decision.

In cell C36, describe the unit of analysis (for example, an individual, a group of students, or another unit such as a classroom). If the unit of analysis is anything other than an individual, provide a description of the unit of analysis and how samples were formed.

Example:

(Yes) Three students participated in the study, each with their own reversal-withdrawal designs, across three different outcomes.

Determine if Outcomes Meet WWC Requirements

Complete **Yellow** and **Orange** Sections of the Data Tab.

Click on cell A39 to navigate to the beginning of the *Data tab*. Complete the yellow and orange section (instructions begin on **page 40**).

- **For all participants, did the authors collect inter-assessor agreement (IAA) for each eligible outcome at least once in each phase?** (Row 40)

In cell C40, use the drop-down menu to select “Yes” if the study provides evidence that the authors collected IAA for each outcome variable and each case at least once in each phase.

If it is clear that the authors did not collect IAA for each outcome variable and each case at least once in each phase, select “No.”

If it is not clear if this requirement was met, select “Uncertain” and note in cell D40 the need for an author query.

In cell D40, briefly describe the evidence for the response. Draft any questions you have on the *Author Query & Response Tab* (instructions begin on **page 48**).

If this IAA information is not provided clearly in a study *and* you send an author query and there is no response, the final MRG should report that the study does not demonstrate that outcomes meet minimum IAA requirements.

Note: If a study uses a statement such as, “IAA data were collected for this outcome across all phases and participants,” reviewers can give the study the benefit of the doubt and assume that IAA data were collected during each phase, for each case, for the outcome.

Examples:

(Yes) *IAA was collected at least once in each phase for each case and outcome.*

(No) *IAA was collected only in the first baseline phase for each outcome. This study does not meet minimum IAA requirements.*

(Uncertain) *IAA was collected for each case and outcome, but it is not clear if IAA was collected at least once in each phase. An author query is required to determine if this requirement was met. (This explanation would be appropriate during an initial review, before reconciliation.)*

(No) *IAA was collected for each case and outcome, but it is not clear if IAA was collected at least once in each phase. An author query was sent to determine if this requirement was met. The author did not respond, so this study does not demonstrate that it meets minimum IAA requirements. (This explanation would be appropriate to use when finalizing the MRG, after the author query time frame has passed.)*

- **Did the authors collect IAA on at least 20% of all sessions?** (row 41)

In cell C41, use the drop-down menu to select “Yes” if the study provides evidence that IAA data were collected on at least 20 percent of all sessions.

If it is clear that this requirement is not met, select “No.”

If it is not clear that this requirement is met, select “Uncertain” and note in cell D41 the need for an author query.

In cell D41, briefly describe the evidence for the response. Draft any questions you have on the *Author Query & Response Tab*.

If this IAA information is not provided clearly in a study *and* you send an author query and there is no response, the MRG should report that the study does not demonstrate that the outcome(s) meet minimum IAA requirements.

Examples:

(Yes) *The study indicates that IAA was assessed during 37 percent of the sessions.*

(No) *The study indicates that IAA was assessed during 18 percent of the sessions. This study does not meet minimum IAA requirements.*

(Uncertain) *The authors did not indicate if they collected IAA on at least 20 percent of all sessions. An author query is required to determine if this requirement was met. (This explanation would be appropriate during an initial review, before reconciliation.)*

(No) The authors did not indicate if they collected IAA on at least 20 percent of all sessions. An author query was sent to determine if this requirement was met. The author did not respond, so this study does not demonstrate that it meets minimum IAA requirements. (This explanation would be appropriate to use when finalizing the MRG, after the author query time frame has passed.)

- **Did the authors collect IAA in at least 20% of the sessions in each condition? (No author query should ask about this issue. A footnote in the report will document whether the study indicated that IAA was collected on at least 20 percent of the data points in each condition.)** (row 42)

In cell C42, use the drop-down menu to select “Yes” if the study provides evidence that IAA data were collected on at least 20 percent of all sessions in each condition.

If the study provides evidence that this requirement is not met, select “No.”

If the study does not provide evidence either way, select “Uncertain.” An author query is necessary only if the study does not indicate whether IAA was collected on at least 20 percent of *all* sessions. A footnote in the report will document whether IAA was collected on 20 percent of all sessions, but it is not clear if IAA was collected on at least 20 percent of the sessions in each condition.

Examples:

(Yes) The study indicates that IAA was assessed during 27 percent of the baseline sessions and 23 percent of the intervention sessions.

(No) The study indicates that IAA was assessed during 22 percent of the baseline sessions and 17 percent of the intervention sessions.

(Uncertain) The study indicates that IAA was assessed during 37 percent of the sessions. The authors did not indicate the percentage of sessions in each condition; a footnote in the intervention report will document this.

- **Does IAA meet minimum acceptable values on each relevant outcome variable (e.g., 80% by percentage agreement or 0.60 by Cohen’s kappa or other correlational measures)? Overall agreement for each outcome must meet the minimum requirements. Note how IAA is documented and provide numerical value(s).** (row 43)

In cell C43, use the drop-down menu to select “Yes” if the study provides evidence that IAA data levels meet the requirements above.

If it is clear that this requirement was not met, select “No.” You should report that the study does not demonstrate that it meets minimum IAA thresholds.

If it is not clear whether IAA data levels meet the requirements, select “Uncertain” and note in cell D43 the need for an author query.

In cell D43, briefly describe the evidence for the response, including how IAA was documented and the specific numerical value(s) provided. Draft any questions you have on the *Author Query & Response Tab*.

If this IAA information is not provided clearly in a study *and* you send an author query and there is no response to the IAA question(s), you should report that the study does not demonstrate that it meets minimum IAA thresholds.

Note: The minimum for percentage agreement—regardless of whether the metric is exact agreement or agreement within one—is 80 percent (or .80). The minimum kappa or correlation is .60. IAA must meet these minimum values for each outcome across all phases or cases, but not separately for each case or phase.

Example:

(Yes) Percentage agreement was calculated by dividing the number of agreements by the sum of agreements and disagreements. IAA for cued vocalizations was 84 percent occurrence (range = 71 -97 percent) and 99 percent nonoccurrence (range=98-100 percent). For spontaneous vocalizations, IAA was 67 percent occurrence (range = 53-67 percent) and 99 percent nonoccurrence (range = 97-100 percent).

- **Is there at least one eligible outcome that meets WWC outcome requirements?** (row 44)

Select “Yes” or “No” in column C (cell C44). List all outcomes that meet WWC requirements with a brief description of the measure (cell D44), including the domain under which the outcome falls. Draft any questions you have on the *Author Query & Response Tab*.

Examples:

(Yes) *Both the listening comprehension and engagement outcomes meet review requirements. These outcomes fall in the reading comprehension and school engagement domains, respectively.*

(No) *The study has one eligible outcome, disruptive behavior, which falls in the problem behavior domain. The authors did not indicate if they collected IAA on at least 20 percent of all sessions and did not respond to an author query, so this study does not demonstrate that it meets minimum IAA requirements. Thus, this study does not have at least one eligible outcome that meets WWC outcome requirements. (This explanation would be appropriate to use when finalizing the MRG, after the author query time frame has passed.)*

Determine if the Independent Variable was Actively Manipulated

- **Is the independent variable (i.e., the intervention) systematically manipulated, with the researcher determining when and how the independent variable conditions change? (Select “No” only if information in the study indicates that the researcher was not directly involved.) (row 47)**

In cell C47, use the drop-down menu to select “Yes” if the study provides evidence that the researcher actively manipulated the independent variable. Manipulation of the independent variable is usually either described explicitly in the Method section of the study or inferred from the discussion of the data collection or results.

If it is clear that the researcher did not actively manipulate the independent variable (for example, the study was conducted retrospectively using existing data), select “No.”

If any language in the study implies that the researcher did not actively manipulate the independent variable, but it is unclear, select “Uncertain” and note in cell D47 the need for an author query.

In cell D47, briefly describe how the timing and implementation of the independent variable were determined. Draft any author questions you have on the *Author Query & Response Tab*.

Examples:

(Yes) The researcher determined in advance that the introduction of the intervention would be staggered across students. Baseline probes were collected for all students. When baseline data were stable for all students, the intervention began with the first student.

(No) Researchers were not involved in the decision to introduce the intervention or reintroduce the baseline condition. Teachers collected data before researchers became involved in the study, and the teachers made these decisions based on the students' behavior.

Determine if there are any Residual Effects between Treatments

- **If there are more than two treatments in the experiment, is the design free from concerns about residual treatment effects, given the specific interventions and outcome(s) in the study? (row 49)**

In cell C49, use the drop-down menu to select “Yes” if the design is free from concerns about residual effects between treatments (that is, there are not likely to be residual treatment effects).

If there are concerns about residual effects among the interventions, select “No.”

Examples:

(Yes) There are three treatments in the experiments for all three cases in the study, but the team determined (in consultation with the content expert) that there are not likely to be residual treatment effects, given the intervention and outcomes.

(No) *Tony’s experiment uses an ABCABC design in which intervention B always comes before intervention C, the intervention under review. In consultation with the content expert, the team determined that intervention B is likely to produce residual effects.*

Attempts to Demonstrate an Intervention Effect

- **Does the design include at least three attempts to demonstrate an intervention effect at three different points in time?** (row 51)

Attempts to demonstrate effects occur when transitioning between conditions.

In cell C51, use the drop-down menu to select “Yes” if there are at least three attempts to demonstrate an effect at three different points in time (for at least one SCD experiment in the study).

If it is clear that this requirement was not met for any SCD experiment, select “No.”

If it is not clear whether the any experiment attempts to demonstrate an effect at three different points in time, select “Uncertain” and note in cell D51 the need for an author query.

In cell D51, briefly describe the evidence for the response. Draft any questions you have on the *Author Query & Response Tab*.

Examples:

(Yes) *There are three reversal/withdrawal designs in this study. The authors used an ABA design for Skip, so there are not three opportunities to demonstrate an effect. Joy and Jen had ABAB designs, which provided at least three attempts to demonstrate an intervention effect at three different points in time.*

(No) *The authors used a multiple probe design across four students. The intervention was introduced to two students at a time—the first two students started the intervention after four baseline sessions and the second pair started after six baseline sessions. Though the study included four students with a total of eight phases, there were not three attempts to demonstrate an effect at three different points in time.*

- **Are there a sufficient number of data points (in each phase)?** Rows 52 through 55 differ based on design—the text in each row reflects the number of data points required for a design to meet standards with or without reservations. Note that review protocols may specify exceptions to the required number of data points. For example, extreme self-injurious behavior might warrant a lower threshold of only one or two data points.
- **Reversal/withdrawal designs** must have at least 5 data points per phase to Meet WWC Single-Case Design Standards without Reservations or 3 or 4 data points per phase to Meet WWC Single-Case Design Standards with Reservations. If not, are the protocol-specific requirements met?
- **Multiple baseline and multiple probe designs** must have at least 5 data points per phase to Meet WWC Single-Case Design Standards without Reservations or 3 or 4 data points per phase to Meet WWC Single-Case Design Standards with Reservations. If not, are the protocol-specific requirements met?
- **Alternating treatment designs** must have at least 5 data points per condition and at most 2 consecutive data points per condition to Meet WWC Single-Case Design Standards without Reservations. They must have at least 4 data points per condition and at most 2 consecutive data points per condition to Meet WWC Single-Case Design Standards with Reservations. If not, are the protocol-specific requirements met? Designs including multiple intervention comparisons (e.g., A vs. B, A vs. C, C vs. B) are rated separately for each comparison.
- **Changing criterion designs** must have at least 5 data points per phase to Meet WWC Single-Case Design Standards without Reservations or 3 or 4 data points per phase to Meet WWC Single-Case Design Standards with Reservations. If not, are the protocol-specific requirements met?

For each design, select “Yes,” “No,” or “Uncertain” to answer the questions in column B. If there is more than one SCD experiment for any type of design, provide information for each SCD experiment separately. If none of the experiments in the study use a particular design, select “Not Applicable” from the drop-down in column C (if you selected “No” for any design(s) in cells C30 through C34, the

options in the corresponding drop-down lists in cells C52 through C55 will be limited to this response).

In column D, briefly describe the evidence for the response. Draft any questions you have on the *Author Query & Response Tab*.

Examples:

(Yes) There are at least five data points in each of the phases for Student 1 and Student 3 (this is sufficient to meet standards without reservations). For Student 2, the first baseline phase and intervention phase have four data points, but the second baseline phase has only two data points (this design does not meet standards).

(Yes) The study uses a multiple baseline design with six phases. Four of the phases contain at least five data points; the other two phases have only four data points. Therefore, the study can only meet standards with reservations.

(Yes) The study uses an alternating treatment design with two interventions. Phases last at most two data points before alternating to the other intervention. There are at least five data points for each intervention (this is sufficient to meet standards without reservations).

Additional Criteria for Multiple Baseline and Multiple Probe Designs

The following criteria apply only to multiple baseline and multiple probe designs. Multiple baseline and multiple probe studies must meet these criteria to *Meet WWC Pilot Single-Case Design Standards Without Reservations* or *Meet WWC Pilot Single-Case Design Standards with Reservations*.

- **If there are no multiple baseline and multiple probe designs in the study, select “Not Applicable” (if you selected “No” in cells C31 and C32, the options in this drop-down list will be limited to this response).**
- **Did all cases have baseline data before the intervention was administered to the first case (i.e., overlapping baselines)? There must be at least three cases with concurrent baseline phases. (row 58)**

If the study meets these conditions, select “Yes” in cell C58.

If it is clear that this requirement is not met, select “No.”

If it is not clear if the requirement is met, select “Uncertain” and note in cell D58 the need for an author query.

If there are no multiple baseline and multiple probe designs in the study, select “Not Applicable” (if you selected “No” in cells C31 and C32, the options in this drop-down list will be limited to this response).

In cell D58, briefly describe the evidence for the response. Draft any questions you have on the *Author Query & Response Tab*.

Examples:

(Yes) All three cases overlapped—the second and third students had at least two baseline data points before the intervention was introduced for the first. The third student had five baseline data points before the intervention was introduced for the second student and three baseline data points after the intervention was introduced for the second student.

(No) Although the sessions in the graphs appear to be concurrent, the authors stated that they used a nonconcurrent multiple baseline design.

- **Do graphs present the timing of sessions consistently across all cases in the design (e.g., does Session 1 for all cases occur at the same time)? (row 59)**

If the graph for at least one experiment presents the timing of sessions consistently across all cases, select “Yes” in cell C59.

If it is clear that this requirement is not met, select “No.”

If it is not clear if the requirement is met, select “Uncertain” and note in Cell D59 the need for an author query.

If there are no multiple baseline and multiple probe designs in the study, select “Not Applicable” (if you selected “No” in cells C31 and C32, the options in this drop-down list will be limited to this response).

Examples:

(Yes) *This is a multiple probe design and the graphs clearly indicate when the sessions for Student 1, Student 2, and Student 3 occurred on the same day. If all students had a probe in the first three sessions, then the graphs clearly had probes when each student had a session and gaps in data when they did not.*

(Unclear) *This appears to be a multiple baseline design, but it is not clear when the sessions for each child occurred. The sessions for Mark, Kim, and Scott are labelled Session 1, Session 2, Session 3, and so on, with no gaps in the data, but they appear to occur at different times (for example, Scott's Session 3 appears to be at the same time as Kim's Session 6).*

- **If the design did not collect data during an intervention or training phase, is the design appropriate given the relevant research question? This design is appropriate if the intervention is expected to produce an effect when all of the intervention components have been implemented; it is inappropriate if the intervention is expected to produce immediate effects when the intervention is implemented. (row 60)**

In cell C60, use the drop-down menu to select “Yes” if the design did not collect data during the intervention phase and this is appropriate given the relevant research question.

Select “No” if the design did not collect data during the intervention phase and this is not appropriate given the relevant research question. The study should be rated *Does Not Meet WWC Pilot SCD Standards* because there are insufficient data to evaluate the attempts to demonstrate an intervention effect.

If the design did not collect data during the intervention phases and it is not clear whether this design is appropriate, select “Uncertain.”

If the design did collect data during the intervention phase, select “Not Applicable.”

If there are no multiple baseline and multiple probe designs in the study, select “Not Applicable” (if you selected “No” in cells C31 and C32, the options in this drop-down list will be limited to this response).

Examples:

(Yes) *The intervention is Self-Regulated Strategy Development, which involves multiple components. The authors do not expect to see changes in students' writing skills until after they have mastered all of the skills in the intervention.*

(No) *The authors did not collect or present data during the intervention phase, but according to the text in the study, they expected the intervention to produce differences in the children's behavior as soon as they introduced the intervention.*

- **Can you determine that the intervention or training phases for each case do not overlap? If so, when any case has a probe after completing training (i.e., the next probe after training or intervention ends), do the other cases still in the baseline phase have continuing baseline measurement at or after that point? (row 61)**

In cell C61, use the drop-down menu to select "Yes" if both requirements are met.

If both of the requirements are not met, select "No."

If the design collected data during the intervention phase, select "Not Applicable."

If there are no multiple baseline and multiple probe designs in the study, select "Not Applicable" (if you selected "No" in cells C31 and C32, the options in this drop-down list will be limited to this response).

Examples:

(Yes) *It is clear from the graph that the intervention phases for all three cases do not overlap. Tom and Mary have baseline probes when Ben has his first probe in the post-intervention phase and Mary has a baseline probe when Tom has his first probe in the post-intervention phase.*

(No) *It is clear from the graph that the intervention phases for Tommy, Betty, and Jerry overlap; they all begin the intervention within one or two sessions of one another.*

(No) *It is not clear from the graph when the intervention phases occur. The graph presents only pre- and post-training data, with no indication of the timing of training sessions.*

Additional Criteria for Multiple Probe Designs

The following criteria apply only to multiple probe designs. Multiple probe studies must meet these criteria to *Meet WWC Pilot Single-Case Design Standards without Reservations* or *Meet WWC Pilot Single-Case Design Standards with Reservations*.

- **Do initial pre-intervention sessions overlap vertically? Within the first three sessions, the design must include three consecutive probe points for each case to *Meet WWC Pilot Single-Case Design Standards without Reservations* or at least one probe point for each case to *Meet WWC Pilot Single-Case Design Standards with Reservations*.** (row 63)

If the study meets either of these conditions, select “Yes” in cell C63.

If it is clear that this requirement was not met, select “No.”

If it is not clear whether either of these conditions is met, select “Uncertain” and note the need for an author query.

If there are no multiple probe designs in the study, select “Not Applicable” (if you selected “No” in cell C32, the options in this drop-down list will be limited to this response).

In cell D63, briefly describe the evidence for the response. Draft any questions you have on the *Author Query & Response Tab*.

Examples:

(Yes) *The baseline sessions overlap and there are three probe points in the first three sessions for each case.*

(Yes) *There are three probe points in the first three sessions for the first and third cases. There is only one probe point in the first three sessions for the second case. Therefore, the highest rating this study can receive is Meets WWC Pilot Single-Case Design Standards with Reservations.*

(No) *The first probe for the third case does not occur until the fifth session. The study cannot meet WWC Pilot Single-Case Designs as a multiple probe design.*

- **Are probe points available just before introducing the independent variable? Within the three sessions just before introducing the independent variable, the design must include three consecutive probe points for each case to *Meet WWC Pilot Single-Case Design Standards without Reservations* or at least one probe point for each case to *Meet WWC Pilot Single-Case Design Standards with Reservations*.** (row 64)

If the study meets either of these conditions, select “Yes” in cell C64.

If it is clear that this requirement was not met, select “No.”

If it is not clear whether either of these conditions was met, select “Uncertain” and note in cell D64 the need for an author query.

If there are no multiple probe designs in the study, select “Not Applicable” (if you selected “No” in cell C32, the options in this drop-down list will be limited to this response).

In cell D64, briefly describe the evidence for the response. Draft any questions you have on the *Author Query & Response Tab*.

Example:

(Yes) *The data include three probe points just before the introduction of the intervention in two of three cases. The third case has only one probe point just before the introduction of the intervention. The highest rating this study can receive is Meets WWC Pilot Single-Case Design Standards with Reservations.*

- **Does each case not receiving the intervention have a probe point in a session in which another case either (1) first receives the intervention; or (2) reaches the prespecified intervention criterion (i.e., a predetermined performance level, such as three consecutive sessions with at least 80 percent of time on task or three consecutive probes with correct responses)?** (row 65)

If the study meets either of these conditions, select “Yes” in cell C65.

If it is clear that this requirement was not met, select “No.”

If it is not clear whether either of these conditions is met, select “Uncertain” and note in Cell D65 the need for an author query.

If there are no multiple probe designs in the study, select “Not Applicable” (if you selected “No” in cell C32, the options in this drop-down list will be limited to this response).

In cell D65, briefly describe the evidence for the response. Please draft any questions you have on the *Author Query & Response Tab*.

Examples:

(Yes) The second case in the series has a probe point in the session in which the first case receives the intervention. The third case has probe points in the sessions in which the first and second cases receive the intervention.

(No) The third case in the series does not have a probe point in the session in which either the first case receives the intervention or reaches the intervention criterion. This study does not meet WWC pilot SCD standards.

Other Study Design Issues

- **Is the study free of other data or analytic issues that would affect the rating (e.g., confounds)?** (row 68)

In column C (cell C68), select “Yes” if there are no other issues about the analysis or data an earlier entry did not capture. If there are issues, select “No.” Summarize any issues in column D (cell D68).

- **Is any attrition not reflected by a reduction in the number of presented data points (e.g., sample loss within a group when a group mean is presented)?** (row 69)

Select a response using the drop-down menu in cell C69.

If the case is a classroom or other group of students, attrition could affect the study results. Attrition can also occur in studies using individuals as cases if the

individual fails to attend all of the planned sessions. Please use column D (cell D69) to describe any attrition not captured in the data points discussion.

Study Rating and Disposition code(s)

- ***What study rating and disposition code should the citation list report?*** (row 71)
This rating is the highest rating applied to any design(s) in the study and is the overall study rating that the report/citation list will document. (rows 71 and 72)

Select a rating using the drop-down menu in cell C71. You will have to select ratings for each SCD experiment in the study before selecting the overall study rating (see rows 73 and 74).

If you select *Meets WWC Pilot Single-Case Design Standards without (or with) Reservations*, complete Stage 3 for the design(s) rated *Meets WWC Pilot Single-Case Design Standards without (or with) Reservations*.

If all of the designs in the study are rated *Does Not Meet WWC Pilot Single-Case Design Standards*, select *Does Not Meet WWC Pilot Single-Case Design Standards* from the drop-down list. Do not complete Stage 3.

If the overall study rating is *Does Not Meet WWC Pilot Single-Case Design Standards* based on the information provided, select the appropriate disposition code in cell D71 and in cell D72 provide an explanation for the disposition code.

When there are multiple SCD experiments within a publication or study, in the intervention report citation list, we will report the disposition code that applies to the majority of SCD experiments in the study. For example, if there are five SCD experiments in a publication and all five do not meet standards because “the eligible outcomes do not meet WWC requirements” (due to not demonstrating minimum IAA thresholds) and two of the five SCD experiments also have “insufficient data,” the intervention report citation list would code the study as “the eligible outcomes do not meet WWC requirements.”

If there is a disposition code tie (e.g., two of four SCD experiments are coded as “there are insufficient data” and two of four SCD experiments are coded as “the eligible outcomes do not meet WWC requirements” *or* both of these disposition

codes are relevant to all four SCD experiments, the citation list should report the *first applicable code from the list below*. (In this example, the study would be coded as “there are insufficient data.”)

Disposition codes for a *Does Not Meet WWC Pilot SCD Standards* rating:

- 1) ... no graphical representation of evidence is provided.
- 2) ... there are insufficient data.
- 3) ... the eligible outcomes do not meet WWC requirements.
- 4) ... the measures of effectiveness cannot be attributed solely to the intervention.
- 5) ... conditions for manipulation of the independent variable are not determined by the researcher.

In some cases, an author query could provide additional information (about IAA, for example), but you decide *not to query the author* because the study already Does Not Meet Standards for a different reason. In this case, select only the disposition codes that you are positive apply, and do not select codes that would require additional information from an author query. Document missing information in C75.

You *should* select disposition codes related to missing IAA information if you sent an author query and did not receive a response. These outcomes do not meet minimum IAA thresholds or requirements.

The *Does Not Meet WWC Pilot Single-Case Design* Disposition Codes are:

... no graphical illustration of evidence is provided.

... there are insufficient data to evaluate the attempts to demonstrate an intervention effect. This covers the following scenarios: (1) designs that do not provide at least three attempts to demonstrate an intervention effect at three different points in time, (2) an ABAB or multiple baseline design with fewer than three data points in at least three phases, (3) an alternating treatment design with fewer than four replications, (4) nonconcurrent multiple baseline or multiple probe experiments, or (5) experiments that should collect outcome data during training but do not collect data. It also covers experiments that used a multiple probe design and (1) initial pre-intervention sessions did not overlap vertically,

(2) probe points were not conducted just before introducing the independent variable, or (3) after the intervention was introduced to a case, each case not yet receiving the intervention did not have at least one additional baseline probe point.

... **the eligible outcomes do not meet WWC requirements.** This covers scenarios in which (1) outcomes are overlapped with the intervention, (2) outcomes are determined not to be sufficiently valid or reliable, or (3) outcomes do not meet minimum inter-assessor agreement thresholds or requirements.

... **the measures of effectiveness cannot be attributed solely to the intervention.** This covers scenarios in which (1) the effects of the intervention of interest were reported only in combination with other interventions; (2) there was a confounding factor (e.g., one teacher delivered instruction in all baseline phases and another teacher delivered instruction in all intervention phases); or (3) there were residual treatment effects.

... **conditions for manipulation of the independent variable are not determined by the researcher.**

Examples:

The study is rated Meets WWC Pilot Single-Case Design Standards without Reservations. The study used one multiple baseline design experiment with three participants and six phases. The data points per phase ranged from six to nine. The outcomes meet review requirements.

The study is rated Meets WWC Pilot Single-Case Design Standards with Reservations. It included one reversal-withdrawal design experiment (ABAB). The first baseline phase had only four data points; the remaining three phases each had six data points. The outcome meets review requirements.

The study is rated Does Not Meet WWC Pilot Single-Case Design Standards, due to insufficient data. It was a reversal experiment that used an ABA design. It did not have the minimum required number of phases to meet standards.

The study is rated Does Not Meet WWC Pilot Single-Case Design Standards, due to insufficient data. It was a multiple probe experiment with three participants. Though the study had sufficient data points per phase to meet standards, participants 2 and 3 lacked a probe point just before the intervention was introduced to them.

The study is rated Does Not Meet WWC Pilot Single-Case Design Standards, because the eligible outcomes do not meet WWC requirements. There were three separate ABAB experiments in the study (for three separate students), and none of the experiments demonstrated minimum IAA thresholds, following an author query. Two of the designs also had insufficient data, because the baseline phases had only two data points.

The study is rated Does Not Meet WWC Pilot Single-Case Design Standards, due to insufficient data. There were four reversal-withdrawal experiments in the study, for four separate students. Two of the experiments were AB designs (and thus did not provide at least three opportunities to demonstrate an effect). The other two designs collected IAA data only during the intervention phases (and thus did not meet minimum IAA requirements). The intervention report will use the insufficient data disposition code.

- ***What is the rating and disposition code for each SCD experiment in the study, given the current information?*** (rows 73 and 74)

Copy and insert rows 73 and 74 for each SCD experiment in the study, so each SCD experiment has at least one rating or disposition code and explanation. For example, if a study has one reversal-withdrawal experiment for Will that does not meet standards because of “insufficient data” (because it was an ABA design), and a second reversal-withdrawal experiment for Kate that does not meet standards because “no graphical illustration of evidence is provided” you should copy and insert rows 73 and 74 so that you can provide separate ratings, disposition codes, and explanations for Will and Kate (i.e., there are two rows for each). In cell C74, enter information specific to the experiment, including the participant’s name, design, and outcome, so the disposition code(s) and explanation clearly refer to a specific experiment.

If a single SCD experiment has multiple disposition codes, copy and insert row 73 for that experiment so you can select each relevant disposition. For example, if there is a reversal-withdrawal experiment for Sam that does not meet standards because the “eligible outcome does not meet WWC requirements” (due to not demonstrating minimum IAA thresholds) and because it has insufficient data (it is an ABA design), you should copy and insert row 73 for Sam’s experiment, so that you can select both reasons this experiment does not meet standards.

In cell D74, enter the explanation for the disposition code(s) selected in row 73 (and the other copied and inserted rows). Do this separately for each experiment.

If you insert multiple dispositions for one SCD experiment (in row 73), make sure to provide explanations for each of the disposition codes in row 74.

If there is only one experiment in the study, you should still complete rows 73 and 74. You can use text from the overall study rating, but add experiment-specific information in C74. You should also include a note in D72 to indicate that there is only one experiment in the study.

Examples:

(Cell C74) *Will (School Engagement)*

(Cell D74) *Will’s ABA reversal-withdrawal experiment does not meet standards because there are only two opportunities to demonstrate an effect.*

(Cell C74) *Kate (School Engagement)*

(Cell D74) *Kate’s reversal-withdrawal experiment does not meet standards because there is no graphical presentation of the data.*

(Cell C74) *Sam (Problem Behavior)*

(Cell D74) *Sam’s ABA reversal-withdrawal experiment does not meet standards for two reasons. First, there are only two opportunities to demonstrate an effect. Second, the problem behavior outcome does not meet WWC requirements because it does not meet the minimum inter-assessor agreement threshold for percent agreement. (Note, for this experiment, you would copy and insert row 73 so there is space for two disposition codes: (1) ... there are insufficient data to evaluate the attempts to demonstrate an intervention effect and (2)... the eligible outcomes do not meet WWC requirements.)*

- **Document any information needed to complete this review** (row 75)
Summarize additional information needed to complete the review, including how the overall study rating and/or ratings for individual SCDs could change. Summarize all missing information here, even if the study or design does not meet standards for another reason.

If an author query is needed to determine the final rating, indicate what response is needed. Refer to the author query in the appropriate row above and the question(s) drafted on the *Author Query & Response Tab*.

For example, if a study does not clearly show that IAA was collected in each phase, but all other criteria for meeting standards are in place, indicate that an author query will be sent to clarify whether IAA was collected in each phase, but that given the information currently available, the study is rated as Does Not Meet Standards. However, pending an author query, the study might be rated Meets Standards.

If a study does not clearly show that IAA was collected in each phase, and the study uses an AB design that does not meet standards due to insufficient attempts, indicate that an author query will not be sent to clarify whether IAA was collected in each phase, because the study already Does Not Meet Standards. In this case, do not use the disposition code “eligible outcome does not meet WWC requirements” (due to not demonstrating minimum IAA thresholds), because you did not give the authors a chance to provide this information in an author query; instead you would only use the disposition code “insufficient data” (because it is an AB design).

When finalizing the MRG (after the author query period has passed) clearly document what information you received or did not receive, in this cell.

Stage 3: Evidence Rating (if at least one design in the study meets standards) (row 81)

DO NOT COMPLETE THIS STAGE IF ALL SCD EXPERIMENTS IN THE PUBLICATION ARE RATED DOES NOT MEET STANDARDS IN STAGE 2, OR BEGIN THIS STAGE IF IT IS POSSIBLE THAT THE STUDY DOES NOT MEET STANDARDS, PENDING THE RESULT OF AN AUTHOR QUERY.

Complete the **Green** (Visual Analysis) Section of the *Data Tab*.

Click on Cell A79 to navigate to the visual analysis section of the *Data Tab*. Complete the green section (instructions begin on **page 43**), and return to the *Main Tab*.

- **What is the highest evidence rating the study receives?** (row 81)

Select the highest rating (strong, moderate, or no evidence) from the drop-down menu in cell C81. Enter an explanation in cell D81.

- **Do any of the contrasts that meet standards in this study provide strong evidence of a causal relationship?** (Row 82)

In cell C82, select “Yes” if any comparison meeting standards provides strong evidence of a causal relationship; select “No” otherwise.

In cell D82, describe each comparison that provides strong evidence of a causal relationship.

Example:

This multiple baseline study provides strong evidence of effectiveness for the intervention “Phonics Now!” on outcome X. The study used three participants, allowing three opportunities to see effects at three different times. In each case, a clear and immediate change in level and trend indicated basic effects. There were no instances of non-effects.

- **Do any of the contrasts that meet standards in this study provide moderate evidence of a causal relationship?** (Row 83)

In cell C83, select “Yes” if any comparison meeting standards is rated as providing moderate evidence of a causal relationship; select “No” otherwise.

In cell D83, describe each comparison rated as providing moderate evidence of a causal relationship. Your description should include the case and the outcomes.

Example:

This multiple probe study provides moderate evidence of effectiveness for using color-coded lists to help students with severe intellectual disabilities prepare meals for themselves. The study included four participants, allowing four opportunities to see effects at different times. For the first, second, and fourth participants, clear effects were present. However, the third participant showed consistently low performance across both the baseline and intervention phase. With three effects and one non-effect, the study provides moderate evidence of a causal relationship.

- **Do any of the contrasts that meet standards in this study provide no evidence of a causal relationship?** (row 84)

In cell C84, select “Yes” if any comparison meeting standards is rated as providing no evidence of a causal relationship; select “No” otherwise. In column D (cell D84), describe each comparison rated as providing no evidence of a causal relationship. Your description should include the case and the outcomes.

Example:

This reversal study provides no evidence that using audio recordings helped the student with learning disabilities complete algebra problems successfully. The ABAB design provided three opportunities to see the effects of using the recordings. Across all sessions, the participant showed steady improvement, with no immediate change in level associated with the introduction or removal of the intervention.

Stage 4: Study Details (if at least one SCD experiment in the study meets standards with or without reservations)

Stage 4 on the main tab of the SRG provides a broad description of the study design and intervention. **This section should describe only SCD experiments that meet standards with or without reservations.** Ideally, this section should be written so that the text for each subsection can be directly pasted into the appropriate appendices for an intervention report or single study review. As such, do not use any text copied directly from the study in the descriptions of the study details in Stage 4.

In summary, describe the following: (row 88)

- **Setting of the study (for example, location, classrooms, courses, schools).** (row 89)

Describe the location of each SCD experiment.

Example:

The study was conducted in an elementary school in a suburban area in the northwestern United States. The intervention for Ben took place in his regular classroom. The intervention for Leo took place in the hallway, right outside of his regular classroom.

- **Study design.** (row 90)

Summarize the study design, including how the introduction of the independent variable was determined.

Example:

This study used a reversal design to identify the effect of an intervention designed to increase on-task behavior. The ABAB design enabled the researcher to observe the student under baseline conditions before introducing the intervention. The intervention was introduced after five sessions of stable baseline performance were observed. It was then withdrawn after eight sessions. It was reintroduced after an additional eight sessions in the second baseline phase.

- **Sample sizes (e.g., students, classrooms, teachers, schools).** (row 91)

Summarize the participants who are eligible and have designs that meet standards (with or without reservations). Report if there are any cases or students that are ineligible for review or do not meet WWC SCD standards.

Example:

Four eligible students participated in the study. One additional student participated in the study, but was ineligible for review.

- **Sample characteristics in protocol (e.g., race, gender, free/reduced-price lunch).** (row 92)

Describe the characteristics of participants who are eligible and have designs that meet standards (with or without reservations). Include any pseudonyms used in the publication. Include racial and ethnic information for the sample, if available.

Example:

James was a 10-year-old white male diagnosed with a learning disability in reading as well as attention deficit/hyperactivity disorder (ADHD). Jack was an 11-year-old Hispanic male whose primary language was Spanish. He was diagnosed with learning disabilities in math, reading, and language arts. Adam was an 11-year-old Hispanic male whose primary language was Spanish. He was diagnosed with learning disabilities in math, reading, and language arts, and with ADHD. Clare was an 11-year-old African American female diagnosed with learning disabilities in math, reading, and language arts.

- **Intervention condition as implemented in the study (including number of days, weeks, or months; number of sessions; and time per session).** (row 93)

Summarize the intervention(s) in sufficient detail to help readers understand what makes this intervention similar to or different from others. The level of detail should be similar to what an introductory section of a typical impact evaluation report would provide. It should include the length of the intervention and the dosage, as well as information about the content and delivery of the intervention. Note any intervention differences between cases or settings. This description should be about the intervention as used in this study, not as described by the developer or in ideal conditions.

Example:

During the intervention phases, students used a checklist to prompt each step of the meal preparation process. The instructor started each session by showing the student a picture of the meal to be prepared and instructed the student to use the checklist to aid in the preparation process. The student then completed each step, checking steps off as they were completed. If the student failed to complete a step correctly, the instructor prompted the student to refer to the checklist. If the student still failed to complete the step correctly, the instructor completed the step outside of the student's view. The instructor then prompted the student to proceed to the next step. Intervention sessions continued until the student completed all steps correctly without having to repeat any of them.

- **Baseline condition as implemented in the study (including number of days, weeks, or months; number of sessions; and time per session).** (row 94)

Indicate procedures used in the baseline (or comparison) condition. Clarify whether the comparison condition was a particular alternative intervention and, if so, name the intervention and provide a brief description if the study provides that information.

Example:

The study used a reversal-withdrawal design. During the baseline/withdrawal condition, there was no formal behavior management system in place in the classroom. Observations revealed that the teacher occasionally praised students or called out to the class that they were doing well. Consequences for inappropriate behavior were inconsistent.

- **Any maintenance phases (include details about timing, duration, data patterns).** (row 95)

Describe any maintenance phases (after the intervention has ended), including the timing, condition, and data trends.

Example:

For Aaron, maintenance data were collected seven weeks after intervention completion. At that time, Aaron's teacher had discontinued the use of his checklist and did not give him opportunities to choose peers for group work. There was a downward trend in Aaron's compliance, and it was at a level similar to the withdrawal phase.

- **All eligible outcomes reported and how they were measured.** (row 96)

Describe each eligible outcome, based on the criteria specified in the protocol under *Types of Outcomes to be Included*. If outcomes vary by case or SCD experiment, specify which outcomes go with each case or SCD experiment. Indicate the domain under which each outcome falls, how each eligible outcome was collected and measured (if it is not self-explanatory from the name of the outcome), and how to interpret each outcome (i.e., the scale of the measure—again, if not self-explanatory). Use quotes around any text that comes directly from the original study.

Example:

There are two eligible outcomes:

Inappropriate behavior (problem behavior domain) was defined as the frequency of inappropriate behaviors during a 10- to 15-minute session, using 10-second partial interval recording. Todd's inappropriate behavior was defined as "screaming, throwing, hitting, crying, yelling, and refusing to move." Mary's inappropriate behavior was defined as "forcefully hitting others." Eli's inappropriate behavior was defined as "leaving the classroom, turning over furniture, and physical aggression toward staff."

Time on task (school engagement domain) was defined as the percentage of time the student was on task and engaged in seatwork, during a 10- to 15-minute session. This outcome was defined similarly for all three students.

- **Any outcomes that are ineligible. Provide details about the outcomes and why they are not eligible.** (row 97)

Outcomes that do not fall into any of the domains in the review protocol are not relevant and not eligible for review. List the outcomes that were not eligible for review along with a reason why. If outcomes vary by case or SCD experiment, specify which outcomes go with each case or SCD experiment.

Example:

The social cooperation outcome is not eligible for review under the Students with Learning Disabilities protocol. It does not fall under any of the topic area's academic domains.

- **Any outcomes that do not meet WWC standards. Provide details about the outcome(s) and a brief rationale for why these outcomes did not meet standards.** (row 98)

List any eligible outcomes that do not meet WWC standards, along with a reason why. If outcomes vary by case or SCD experiment, specify which outcomes go with each case or SCD experiment.

Example:

The study did not present data for the "on-task behavior" outcome visually, so this outcome does not meet WWC standards.

- **Support for implementation.** (row 99)

Indicate both the staff training and technical assistance conducted to support the implementation of the intervention (as evaluated in the study).

Example:

The students' regular classroom teacher conducted the instruction in both the baseline and intervention conditions. He met privately with the researcher before the start of the study and received brief training on how to use the instructional videos as part of the intervention.

- **Implementation fidelity (how was it assessed and level obtained).** (row 100)

Describe the steps the researcher took to ensure fidelity in baseline and intervention procedures.

Example:

The instructor followed a detailed script when implementing both the baseline and intervention procedures. Implementation fidelity was measured simultaneously with IAA. Independent observers determined that baseline and intervention procedures were implemented correctly in 100% of sessions for both participants.

- **Discuss other SCD experiments presented in the study but that were not part of the review. Briefly describe the reason this review focuses on a specific experiment(s).** (row 101)

Describe any additional SCD experiments presented, but not already covered or described in the current SRG. Explain why the review did not include these SCD experiments.

Example:

The authors also presented an SCD experiment used to measure the effect of a different intervention, called Patch. This SCD experiment is not relevant for the current review effort, because it focuses on Functional Behavioral Assessment, and not Patch.

Data Tab of the SRG

Each row corresponds to a research question addressed by an SCD. Represent each SCD experiment (with an eligible outcome) on a separate row. If a study contains data on more than one outcome for a participant, each outcome should receive a separate row.

For example, use a separate row to represent each reversal-withdrawal design presented in a study. If there are three separate reversal-withdrawal designs for three participants, each participant would get his or her own row. If there are separate reversal-withdrawal designs for three different outcomes (across one participant), each outcome would get its own row.

Represent multiple baseline and multiple probe designs on one row, such that the row represents the entire SCD experiment. Again, if the multiple baseline or multiple probe design offers data on more than one outcome, each outcome would get a separate row.

For alternating treatment designs, each comparison (that is deemed relevant to the review) should have its own row.

Yellow (Data) and Orange (Outcomes) Sections of the Data Tab

General instructions:

- Include only outcomes that fall within one of the eligible domains for the review.
- If you are uncertain whether a particular outcome measure falls under one of these domains, check for additional guidance that your team can provide on outcome measures.

- **Measure.** (column A)
Enter the name of the outcome or test, exactly as specified in the article or report.

- **Domain.** (column C)
Indicate the eligible domain(s) from which the outcome comes. This should be the name of the *domain only*, no constructs.

- **Construct.** (column D)
Note any relevant constructs for the domain in this column. Check the protocol for details.

- **Design.** (column E)
Provide the SCD type.
- **Sample.** (column F)
List the participants (sample size) for the experiment in this column (for example, enter “Maggie [1]” for a reversal/withdrawal experiment focused on one participant; “Low-performer Group [10]” for a reversal/withdrawal experiment focused on a group of 10 low-performing students; or “Steve, Michael, and Adam [3]” for a multiple baseline experiment across three participants).
- **Outcome period.** (column G)
Indicate any relevant information about the timing of intervention or measures (e.g., baseline and intervention).
- **Binary.** (column I)
Select “Yes” from the drop-down menu if the outcome is a 1/0 variable for which the underlying construct is a yes/no answer, such as “ever graduated” or “retained in grade” (i.e., there is not an underlying distribution of the variable). Select “No” otherwise.
- **Standardized test.** (column J)
Select “Yes” from the drop-down menu if the test is a standardized test. The score should be from the full test or established subscale to be considered a standardized test.
- **Face validity.** (column K)
Select “Yes” from the drop-down menu if the measure appears to be a reasonable assessment of the construct it purports to measure; select “No” if you see an obvious problem with the measure, and indicate your concerns in column P.

- **Not overaligned?** (column L)

Select “True” from the drop-down menu if you have no concerns about the measure being overaligned with the intervention; select “False” if you have concerns about the measure being overaligned with the intervention.

Explain any concerns in column P.

- **Consistent measure and data collection process?** (column M)

Select “Yes” from the drop-down menu if the intervention and baseline conditions collected the same measure in a similar manner; select “No” if it is clear that the intervention and baseline conditions collected outcome data in a different manner, potentially in a way that could lead to differences in average outcomes by condition.

Explain any concerns in column P.

- **Meets IAA requirements (see Main Tab)?** (column N)

Select “Yes” from the drop-down menu if the IAA requirements are met (Main Tab Rows 38 to 41). If the study does not demonstrate that the requirements are met, select “No” from the drop-down menu.

Explain any concerns in column P.

- **Meets WWC requirements?** (column O)

Based on the answers in the previous columns and the standards established in the review protocol, select “Yes” from the drop-down menu if the outcome meets all of the requirements; select “No” otherwise.

If you select “No,” ensure that column P includes details on why the measure does not meet WWC requirements. (Note: The measure name in column A will turn red to indicate it should not be reported.)

- **Notes or concerns about the measure.** (column P)

Summarize any concerns you have about this measure. In particular, if the measure does not meet the requirements for inclusion in this review, indicate why.

Now, go back and enter a decision in cell C42 on the *Main Tab* (Is there at least one relevant outcome that meets review requirements?).

Green (Visual Analysis) Section of the *Data Tab*

For each outcome, complete the first step of the visual analysis:

- **Baseline data pattern.** (columns R and S)

Do the data in the first baseline phase for each case document that (1) the proposed concern is demonstrated and (2) there is sufficient demonstration of a clearly defined baseline pattern that can be used to assess the effects of an intervention? For alternating treatment designs, the data must document only that the proposed concern is demonstrated. In experiments with no baseline phase, the proposed concern can be demonstrated through (1) a business as usual condition, (2) a description of the problem in the text, or (3) a determination from the lead methodologist that the experiment does not have to demonstrate a concern because of ethical concerns.

If the outcome data meet both of these conditions, select “Yes” in column R. If not, select “No.”

Describe the data pattern or other evidence (alternating treatment designs only) that supports your answer in “Supporting Information, Concerns, or Questions” (column S).

If outcome data do not meet these conditions, skip to the “Rate the Evidence” step below (page 45), select “No Evidence” in column AC and enter “n.a.” in columns AF, AG, and AH under demonstrations of an effect, demonstrations of a non-effect, and attempts to demonstrate an effect.

If the outcome data did meet these conditions, complete the remaining steps of the visual analysis. If the baseline data for any tier in a multiple baseline or multiple probe design meet these conditions, complete the remaining steps of the visual analysis. This is necessary to accurately determine the number of demonstrations of an effect and demonstration of a non-effect.

Examples:

(Yes) *Performance during baseline was consistently low, with no more than 20% of steps completed correctly in any session.*

(No) *The student demonstrates a very low error rate (though with some variability) in the baseline phase. There is no clear need for intervention.*

- **Compare level, trend, and variability for adjacent phases.** (column U)

Are there differences in the level, trend, and variability of data between each phase and the next phase? (For alternating treatment designs, the trend does not need to be assessed due to the limited number of data points in each phase.) Describe the data patterns in column U.

Examples:

(Yes) *There is a substantial increase in the level of the data points, and an upward trend during each intervention phase. Variability was low in both conditions.*

(No) *There are differences in the trends (downward during baseline and upward during intervention). However, there is no clear change in levels between the phases, and variability is such that the data range from 0 to 100% in both conditions.*

- **Compare overlap and immediacy of effect for adjacent phases.** (column W)

Is there an immediate effect and an absence of overlap between each phase and the next phase? Describe the data patterns in column W.

Examples:

(Yes) *There were immediate effects with no overlap between the baseline and intervention phases.*

(No) *The performance in the intervention phase was higher than in the baseline phase, on average; however, the increase was not immediate. There was also substantial overlap between the phases, despite the change in level.*

- **Compare consistency of data patterns (i.e., level, trend, and variability) within phases of the same condition.** (column Y)

Are the data patterns in phases of the same condition consistent? (For alternating treatment designs, the trend does not need to be assessed due to the limited number of data points in each phase.) Describe the data patterns in column Y.

Examples:

(Yes) All baseline phases show low performance with low variability. Intervention phases show substantially increased levels and high variability.

(No) The two baseline phases show similar low levels of performance with low variability. However, the two intervention phases are not very similar. The first shows a moderate level with high variability and a stable trend. The second shows a much higher level with moderate variability and an upward trend.

For multiple baseline and multiple probe designs, complete the visual analysis step described in column AA. For all other designs, skip to the “Rate the Evidence” step described in column AC.

- **For Multiple Baseline and Multiple Probe Designs Only:**

Vertically compare each baseline series to determine if changes in data patterns in each tier are independent of one another. When one case either first receives the intervention or reaches a prespecified condition, are the probes for other cases not receiving the intervention consistent in level and trend with their previous baseline points? Is the time in which an intervention effect is initially demonstrated with one series (e.g., the first five days following introduction of the intervention for participant 1) independent of changes in the data pattern over the same time frame in the other series of the design? (column AA)

Describe the data patterns in column AA. If the study does not use a multiple baseline or multiple probe design, type “n.a.”

Examples:

(Yes) There were no changes in baseline patterns for any cases when other cases started the intervention or had data indicating there was an intervention effect.

(No) Participants 2 and 3 showed an increase in performance in session 6, when the intervention was introduced for Participant 1. Participant 3 also showed a clear increase in performance in session 11, when the intervention was introduced for Participant 2.

For alternating treatment designs, complete the visual analysis step described in column AB. For all other designs, skip to the “Rate the Evidence Step” described in column AC.

- **For Alternating Treatment Designs Only:** Compare the overall mean levels for the intervention and comparison conditions. The overall mean should include all points, including outliers, for each condition. Do they clearly demonstrate a visual effect? (column AB).

Describe the mean levels in column AB. If the study does not use an alternating treatment design, type “n.a.”

Examples:

(Yes) The overall mean for the intervention is well above the overall mean for the comparison (all of the intervention probes are at a higher level than the comparison probes).

(No) The overall means for the intervention and comparison sessions are very similar. There are a large number of data points, so even though there are more than three demonstrations of an intervention effect, the overall means do not clearly demonstrate that the intervention had an effect on time on task.

- **Characterize the evidence.** (columns AC and AD)

Are there at least three effects at three different points in time (providing evidence of a causal relationship)? The possible levels of evidence are as follows:

- Strong evidence: At least three effects and no non-effects
- Moderate evidence: At least three effects, but also either (1) at least one example of a non-effect or (2) multiple baseline data or multiple probe data suggesting that an effect demonstrated in one series is associated with change in the data pattern for another series

- No evidence: (1) Fewer than three effects at different points in time or (2) the baseline data do not demonstrate the proposed concern
 - Alternating treatment designs can only show strong or moderate evidence of a causal effect if there are no clear effects in the opposite direction and the overall means clearly demonstrate a visual effect. If there are clear effects in the opposite direction or the overall means do not clearly demonstrate a visual effect, there is no evidence of a causal effect.

Select the level of evidence observed in column AC. Describe the data patterns that support your answer and any additional notes, concerns, or questions you have about the evidence in “Supporting Information, Concerns, or Questions” (column AD).

Examples:

(Strong evidence) This is an ABAB study with three clear effects. In all three phase changes, the observed pattern of data after the phase change deviates from the projected pattern before the phase change. There were immediate increases in level and trend. There was also minimal overlap in data points between baseline and intervention phases.

(Moderate evidence) This multiple baseline design across four participants shows three clear effects (for Eli, Isaac, and Sid), and one non-effect (for Milo). For Eli, Isaac, and Sid, there were clear increases in level and trend following the introduction of the intervention. For Milo, there was no change in level from the baseline to the intervention phase.

(No evidence) This multiple baseline study does not provide evidence of effectiveness. Though all three participants show changes in level associated with the intervention, there is a clear upward trend throughout both the baseline and intervention phases in all series. There is also overlap in data points between baseline and intervention conditions, resulting in no demonstrations of intervention effects.

- **Summarize the WWC Findings:** Provide the total number of demonstrations of an effect, demonstrations of a non-effect, demonstrations of an effect in the opposite direction, and attempts to demonstrate an effect for each experiment, participant, or outcome.

For each SCD experiment, enter the total number of demonstrations of an effect in column AF, the total number of demonstrations of a non-effect in column AG, the total number of demonstrations of an effect in the opposite direction in column AH, and the total number of attempts to demonstrate an effect in column AI.

For alternating treatment designs only: Ambiguous effects should not be counted as demonstrations of an effect (in column AF) or demonstrations of an effect in the opposite direction (in column AH).

- Ambiguous effects in the *same direction* as the clear effects in column AF **will not be included in the counts** in columns AF, AG, or AH, but should be described in column AD.
- Ambiguous effects in the *opposite direction* as the clear effects in column AF should be counted as **non-effects** (in column AG).

After completing the visual analysis, go back to the *Main Tab* to select the highest evidence rating in cell C81 and enter any supporting information, concerns, or questions there. Complete the entries in columns C and D in rows 82, 83, and 84.

Author Query and Response Tab of the SRG

Reviewers should draft potential author query questions as part of their review of a study. The MRG should capture the final set of questions as sent to the author. It should also document whether (and when) a response was received and the response.

- **Date author query sent** (cell B3)

Enter the date the author query was sent.

Note: The team coordinator will send author queries from the WWC email account for reviews conducted under contract ED-IES-13-C-0010.

- **Response received? (Yes/No)** (cell B4)

Enter “Yes” or “No” to document whether any response was received.

- **Date of response** (cell B5)

Enter the date the team received the response(s).

- **Mode of response (e.g., fax, email)** (cell B6)

Enter the particular mode(s) by which the team received the response(s).

- **Author query questions and answers**

Save questions and responses in rows 8 through 15. Copy and insert additional rows as needed. If the author query included a table, enter the values the author provided in any table shell (or copy and paste populated tables, if appropriate).

Summary Tab of the SRG

This worksheet captures high-level information about the study, rating, outcomes, and findings.

Many of the cells in this worksheet will populate automatically based on the data entered in the Main Tab and Data Tab of the SRG. These cells are shaded gray. They include summary-level and experiment-specific information:

Study-specific entries

- Study ID (cell A4)
- Full citation (cell B4)
- Review date (cell T4)
- Standards version (cell B6)
- Review protocol and version (cell B6)

Experiment-specific entries (starting in row 10)

- Measure (column A)
- Domain (column B)
- Construct (column C)
- Design (column D)
- Sample (column E)
- Visual analysis rating (column H)
- Demonstrations of an effect (column I)
- Demonstrations of a non-effect (column J)
- Demonstrations of an effect in the opposite direction (column K)
- Attempts to demonstrate an effect (column L)

In addition to the prefilled cells, reviewers will have to add data on the age(s) or grade(s) associated with each case, the WWC design rating for each experiment, and author-reported findings. These cells are shaded green.

- Enter the age(s) or grade(s) associated with the experiment's sample in column F.
- Use the drop-down menu to select the WWC rating associated with each SCD experiment in column G.
- In column T, summarize the author-reported findings for each SCD experiment.

Examples:

The study authors found that pivotal response training decreased Amy's problem behavior.

The study authors reported that the intervention produced a small increase in Jill's on-task behavior when first introduced and a large increase in on-task behavior when the intervention was re-introduced.