



What Works Clearinghouse  
Study Review Guide Instructions for  
Reviewing Regression Discontinuity Designs

Released January 2022

This document provides step-by-step instructions on how to complete the Study Review Guide (SRG, S4.1; January 2022 release) for regression discontinuity designs (RDDs).

For every What Works Clearinghouse (WWC) review, reviewers will need to complete an SRG. 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 Handbooks. At the end of the review process, a Master SRG (MRG) will represent the final assessment of the study and provide a record for the WWC, as well as serve as a key input in producing reports. For more details on the review process, see the instructions provided in the Procedures Handbook or the review protocol.

*This guide is intended for individuals trained and certified in the version 4.1 WWC Standards and Procedures Handbooks. The current versions of the WWC Standards and Procedures Handbooks are available at <https://ies.ed.gov/ncee/wwc/handbooks>; review protocols are available at <https://ies.ed.gov/ncee/wwc/handbooks#protocol>; and the most current version of the SRG is available at <https://ies.ed.gov/ncee/wwc/StudyReviewGuide>.*

# Contents

	Page
General Characteristics of the SRG .....	1
<i>Main Tab</i> of the SRG .....	1
Stage 1: Preliminary screening .....	1
Overview .....	2
Screening criteria .....	2
RDD screening criteria .....	5
Stage 2: Quality of evidence (if the study passes stage 1) .....	7
Design details .....	7
Complete columns A-L of the data tab (row 30) .....	8
Standard 1: Integrity of the forcing variable (row 32) .....	9
Standard 2: Attrition .....	11
Standard 3: Continuity of the relationship between the outcome and the forcing variable .....	12
Standard 4: Functional form and bandwidth .....	14
Standard 5: Fuzzy regression discontinuity design (FRDD) .....	17
Stage 3: Study details .....	21
<i>Data tab</i> of the SRG .....	26
Outcome name, domain, and measure characteristics section of the data tab (columns A-L) .....	26
General instructions .....	26
Characteristics .....	27
Attrition section of the data tab (columns M-AG) .....	28
Baseline equivalence section of the data tab (columns AH-AW) .....	30
Analysis and results section of the data tab (columns AX-BE) .....	32
Findings section of the data tab (columns BF-BP) .....	32
<i>Summary tab</i> of the SRG .....	34
Author query & response tab of the SRG .....	35
Reference .....	36

## GENERAL CHARACTERISTICS OF THE SRG

In the *Main Tab*, a reviewer may need to add rows in two places:

- To represent all appropriate disposition codes for why the study is identified as *Not Eligible for Review* either as a study generally or as an RDD specifically (row 17 and row 24).
- To represent all appropriate disposition codes for why the study is rated *Does Not Meet WWC RDD Standards* (row 69).

To add additional rows to capture the disposition, click on the *Review Tab* and click “Unprotect Sheet.” Place your cursor in the row below (that is, cell D17 and cell D24); from the *Main Tab*, click “Insert” and select “Insert Sheet rows.” Select cells A-D for the row you are copying. Press Ctrl + C. Move your cursor to cell A in the row you added. Press Ctrl + V. On the *Review Tab*, click “Protect Sheet” to ensure that you do not overwrite a formula by mistake.

The *Data Tab* includes 20 rows for outcomes (rows 2-21). If the study you are reviewing has more than 20 outcomes × sample × time period, click on the *Review Tab* and click “Unprotect Sheet.” Highlight a blank row. Press Ctrl + C. Right-click and select “Insert Copied Cells.” Repeat to add as many rows as you need. On the *Review Tab*, click “Protect Sheet” to ensure that 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 and is structured in three separate steps, as described in this section.

Your role as a reviewer is to provide complete information for each element in a given section. In general, you will need to enter data into columns B, C, and D in the *Main Tab*.

- Column B typically will require a short answer to a question posed in column A.
- Column C will require a justification for the short answer in column B. As much as possible, information in column B should be in complete sentences and use plain language, avoiding acronyms whenever possible.
- Column D will require the page numbers from the study that serve as the source of the justification presented in column C.

### Stage 1: Preliminary screening

Stage 1 of the study review assesses whether the given study is eligible for WWC review under a given review protocol as well as whether it meets the RDD screening criteria. All reviews are conducted under a specific review protocol. All screening and review decisions relate to that specific protocol and could differ under another review protocol. For instance, a study that is ineligible under one protocol could be eligible under another protocol because the sample or outcome measures align more closely with the latter than the former. Similarly, a study could meet standards under one protocol but not another because of aspects of the research design that play out differently under different review protocols.

- **Full Citation** (cell A1).
  - Enter the full citation for the study.

- Example: Castleman, B. L., & Long, B. T. (2013). *Looking beyond enrollment: The causal effect of need-based grants on college access, persistence, and graduation* (NBER Working Paper No. w19306). National Bureau of Economic Research.
- **Study ID** (cell A2): Enter the Study ID for the review.
- **Reviewer Number** (cell A3): Enter your reviewer number. If a Master Review Guide, enter “MRG.”
- **Review Date** (cell A4): Enter the date of your review.

## Overview

- **Standards and Protocol** (row 5): Enter the version of standards being used for the review (cell A5). Enter the name and version of the protocol used for the review (cell A5). All WWC reviews take place using the Study Review Protocol, version 4.1.
- **Intervention Name** (row 6).
  - Clearly indicate the name of the intervention(s) you are reviewing within this SRG (cell A6).
  - Note whether a single name refers to multiple versions of an intervention or different names refer to the same or related interventions (cell B6).
  - Example: *Florida Student Access Grant*.

## Screening criteria

- **Effectiveness** (row 8).
  - Does the study examine the effect of an intervention?
  - In cell B8, 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** (row 9).
  - Does the study use an eligible design (randomized controlled trial, quasi-experimental design, RDD, or single-case design)?
  - In cell B9, select “Yes” if the study uses any design with a comparison group or condition, regardless of the quality of the design; select “No” if the study uses a design without a comparison group or condition, such as a pre-post design that lacks a comparison group, a meta-analysis, or a literature review.
  - In column C, enter the type of design. If the design is a comparison group design or a single-case design, stop entering information into the RDD SRG template. Provided that you are certified to conduct a review for the given design, obtain access to the appropriate SRG for the randomized controlled trial/quasi-experimental design or single-case design, and complete the SRG in the correct format. If you are working with the WWC, please route the study in question to review team leadership so they can obtain a reviewer certified in the appropriate design standards.
- **Focus** (row 10).
  - Is the intervention a program, product, policy, or practice with the primary focus aligned with the review protocol?

- In cell B10, select “Yes” if the intervention meets the criteria for inclusion specified in the protocol under *Types of Interventions to Be Included*; select “No” otherwise. Provide details about the intervention in cell C10.
- **Sample Alignment** (row 11).
  - Does the study meet the requirements for sample characteristics specified in the review protocol?
  - In cell B11, select “Yes” if the intervention meets the criteria for inclusion specified in the protocol under *Types of Populations to Be Included* (for example, percentage of English learner students or percentage of general education); select “No” otherwise.
  - In column C, describe the sample with respect to how it satisfies or does not satisfy this requirement. If the study is rated either *Meets WWC RDD Standards Without Reservations* or *Meets WWC RDD Standards With Reservations*, a more detailed description of the study participants will be provided in stage 3 of the SRG.
  - Examples:
    - » *(Yes) The sample includes 60 percent of students with learning disabilities.*
    - » *(Yes) The sample includes teachers working in Washington State public schools that qualify for Challenging Schools Bonus eligibility, for the school years 2001/02-2012/13. The sample includes elementary, middle, and high school teachers. This sample aligns with the sample characteristics specified in the review protocol.*
- **Time** (row 12).
  - Was the study published within the time frame relevant to the review protocol?
  - In cell B12, select “Yes” if the study falls within the time frame outlined in the protocol; select “No” otherwise.
- **Age or Grade Range** (row 13).
  - Does the study examine students in the age or grade range specified in the review protocol?
  - In cell B13, 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 C, please indicate the grade levels of the students in the sample or the age range of the children in the sample, if the sample is described in terms of age rather than grade level. If the study receives a rating of either *Meets WWC RDD Standards Without Reservations* or *Meets WWC RDD Standards With Reservations*, a more detailed description of the study participants will be provided in stage 3 of the SRG.
  - Example: *(Yes) Ninety percent of the intervention students were identified as ninth graders. Ninety-five percent of the comparison students were identified as ninth graders. The mean age of the intervention group is 15.5 years (standard deviation = 1.20). The average age of the comparison group is 15.7 years (standard deviation = 1.15).*
- **Location** (row 14).
  - Does the study examine sample members in a location specified for the review protocol?

- In cell B14, select “Yes” if the study sample was drawn from the geographic region described in the protocol under *Types of Populations to Be Included*; select “No” otherwise.
- In cell C14, describe the location. If the study is rated either *Meets WWC RDD Standards Without Reservations* or *Meets WWC RDD Standards With Reservations*, a more detailed description of the study setting will be provided in stage 3 of the SRG.
- **Outcomes (row 15).**
  - Does the study address at least one outcome in a domain relevant for the review protocol?
  - In cell B15, select “Yes” if the study estimates the impacts of the intervention on at least one outcome that falls into one of the eligible domains specified in the protocol under *Types of Outcomes to Be Included*.
  - In cell C15, summarize the types of outcomes included (you do not need to list each outcome here). For example, you might enter the names of the domains for which there are outcomes and the number of outcomes in each domain.
  - Examples:
    - » *The study reports three math outcomes—full scores, numeracy, and geometry/spatial subscales—from the research-based early mathematics assessment developed by Clements et al. (2008).*
    - » *The study addresses three outcomes in domains relevant to the review protocol: teacher retention in the same school during the following academic year, math achievement test scores, and reading achievement test scores.*
  - Select “No” if no outcomes fall within the domains of the review protocol.
  - In column C, briefly list the outcomes that are available. Provide any additional documentation needed for a reviewer to confirm that the outcomes are not eligible under that protocol. If you are working on a WWC team and are uncertain whether any of the study’s outcomes qualify under the protocol, seek additional guidance from your review leadership.
  - Example: *The study is not eligible for review under the Dropout Prevention review protocol. Although the study is an examination of a youth development program that aims to keep youth engaged in school, this publication only reports outcomes related to sexual behavior (pregnancies, use of birth control including condoms).*
- **Does the study meet WWC and review protocol screening criteria (row 16)?**
  - If more than one disposition is appropriate, copy and paste this row and select the additional disposition code.
  - From the drop-down menu (“Yes/No”) in column B, select “Yes” if the study met all the screening criteria in this section and warrants a full review to determine whether the study design meets WWC standards. In column C, select “The study is eligible for review. An explanation for that decision is below” from the drop-down menu.
  - Select “No” in column B if the study failed to meet any of the screening criteria in this section and select the appropriate screening disposition code in column C.

- The Screening Disposition Codes are as follows:
  - » ... **is out of scope of the protocol**. This covers scenarios in which the study did not use an intervention reported in English, occur within the time frame specified in the protocol, occur within a geographic area specified in the protocol, occur within a setting specified in the protocol, or include an outcome within a domain specified in the protocol.
  - » ... **does not use a sample aligned with the protocol**. This covers scenarios in which the study did not use a sample within the age or grade range specified in the protocol, or [the study did not show/the authors could not confirm/the WWC could not confirm] that at least 50 percent of the sample was classified [review specific classification].
  - » ... **does not use an eligible design**. This covers scenarios in which the study did not examine the effectiveness of the intervention; contain a primary analysis; use a comparison group design, RDD, or single-case design; or provide adequate or consistent information to assess whether it was eligible for review.
- **Explanation for Screening Disposition (row 17).**
  - If you selected “Yes” in cell B16, provide a brief explanation of why the study is eligible for review in cell C17.
  - Example: *This study meets WWC and review protocol screening criteria. The study uses a fuzzy RDD to examine the effects of a teacher certification incentive bonus program on teacher retention and student achievement outcomes. The sample includes teachers providing instruction to students in grades K-12 in Washington state public high schools; the study was published in 2018.*
  - If you selected “No” in cell B16, provide a complete explanation for the screening disposition you selected in cell C17. One option is to consider using the scenarios described earlier as the starting point for your explanation.
  - Example: *The publication is not eligible for review under the Beginning Reading review protocol; the students are in fourth and fifth grades. It may be eligible for review under the Adolescent Literacy review protocol because it is a study of Reading Mastery and reports reading scores from the SAT-11.*

### **RDD screening criteria**

- **Are the groups formed, at least in part, using a numerical forcing variable (row 19)?**
  - The RDD standards require that participants with numbers at or above a cutoff value are assigned to intervention (alternately, at or below a cutoff value), and participants with numbers on the other side of the cutoff are assigned to comparison.
  - In cell B19, select “Yes” if the groups are formed, at least in part, using a numerical forcing variable. Select “No” if the forcing variable is entirely nonnumerical.
  - In cell C19, describe the forcing variable, including the cutoff that participants must be above or below to be eligible for the intervention.
  - Examples:
    - » *(Yes) Expected Family Contribution was the forcing variable. During the 2000/01 school year, students whose expected family contribution was less than \$1,590 were eligible for a \$1,300 Florida*

*Student Access Grant (2000 constant dollars); this translates to families with incomes roughly below \$30,000 that year (\$40,300 in 2011 dollars) being eligible for a Florida Student Access Grant.*

» *(Yes) Proportion of free or reduced-price lunch is the forcing variable used to determine eligibility for the Challenging Schools Bonus. Eligibility is based on two sources of free or reduced-price lunch enrollment data: the Child Nutrition Services school administrator-reported survey and the Core Student Record System derived from official enrollment records. Until 2011, schools were grandfathered in for eligibility based on their school's free or reduced-price lunch enrollment from the prior school year. The thresholds used to determine Challenging Schools Bonus eligibility approximate the 75th percentile of school free or reduced-price lunch enrollment, such that schools above the 75th percentile are eligible for the Challenging Schools Bonus program, whereas those below the 75th percentile are not eligible for the program.*

- **Is the forcing variable ordinal and are there at least four unique values above and below the cutoff (row 20)?**
  - The RDD standards require that the forcing variable is ordinal in that it has a unique ordering of values from lowest to highest. Continuous variables, such as income in U.S. dollars, would qualify as ordinal.
  - In cell B20, select “Yes” if the forcing variable is ordinal and there are at least four unique values above and below the cutoff. Select “No” if the forcing variable is not ordinal or if there are fewer than four unique values above or below the cutoff.
  - In cell C20, describe the nature of the forcing variable (for example, nominal, ordinal, continuous) and whether there is a sufficient number of values above and below the cutoff.
  - Example: *(Yes) Expected family contribution was the forcing variable. During the 2000/01 school year, students whose expected family contribution was less than \$1,590 were eligible for a \$1,300 Florida Student Access Grant (2000 constant dollars); this roughly translates to families with incomes below \$30,000 that year (\$40,300 in 2011 dollars) being eligible for a Florida Student Access Grant.*
- **Is it true that the cutoff value was not used to assign members of the study sample to interventions other than the one being tested (row 21)?**
  - If the cutoff was used to assign members of the sample to another intervention, it would represent a confounding factor in the design.
  - In cell B21, select “Yes” if the cutoff value was not used to assign members of the study sample to interventions other than the one tested. Select “No” if the cutoff value was used to assign members of the study sample to an intervention other than the one being tested.
  - In cell C21, describe any important supporting details.
  - Example: *(Yes) There is no indication that the cutoff value was used to assign members of the study sample to other interventions. All students (both those who were and were not eligible for Florida Student Access Grant) were eligible for at least a \$1,750 federal Pell Grant. Therefore, the counterfactual is other need-based aid (such as the Pell Grant), and the study investigates the effects of a Florida Student Access Grant layered on top of other need-based aid (that is, the Pell Grant).*

- **Is the forcing variable that was used to calculate impacts the same as the forcing variable that was used to assign units to groups (row 22)?**
  - Authors will sometimes use as their forcing variable a variable intended to approximate the actual forcing variable. If the authors use anything other than the actual forcing variable, select “No” in cell B22.
  - If the authors use the same forcing variable as was used to assign units to groups, select “Yes” in cell B22.
  - Consult review team leadership if you are uncertain about whether the forcing variable used in the study is the same as was used to assign units to groups.
- **Does the study meet WWC RDD screening criteria (row 23)?**
  - If more than one disposition is appropriate, copy and paste this row and select the additional disposition code.
  - From the drop-down menu (“Yes/No”) in column B, select “Yes” if the study met all the screening criteria in this section and warrants a full review to determine whether the study design meets WWC standards. In column C, select “The study is eligible for review. An explanation for that decision is below.”
  - Select “No” in column B if the study failed to meet any of the screening criteria in this section and select the appropriate screening disposition code in column C.
  - The Screening Disposition Codes are as follows:
    - » **... forcing variable is not numeric.** This covers scenarios in which the study did not use a numeric forcing variable to assign participants to the intervention or comparison groups.
    - » **... forcing variable is not ordinal or there are fewer than four unique values on at least one side of the cutoff value.** This covers scenarios in which the cutoff values are numeric but do not have a unique ordering, or there were fewer than four unique values above and/or below the cutoff.
    - » **... the cutoff value was used to assign members of the study sample to an intervention other than the one being tested.** This covers scenarios in which the cutoff value was used to assign participants to one or more interventions other than the one being tested.
    - » **... the forcing variable used to calculate impacts is not the same as the forcing variable used to assign units to groups.** This covers scenarios in which the study uses a proxy for the forcing variable, such as might be the case if the original forcing variable is not available to the researcher.

## 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 be making assessments about the level of evidence from the study. To improve the transparency of the review process, whenever possible, use direct quotes from the study when asserting that a study does not pass a particular standard.

### *Design details*

- **How are the intervention and comparison groups formed (row 28)?**

- In cell B28, use the drop-down menu to select the appropriate design (RDD or Cluster RDD). The selection of a design results in the highlighting of cells on the *Data Tab* to guide you to the relevant cells for that design for attrition and baseline equivalence calculations.
- In cell C28, describe each stage in the process by which the intervention and comparison samples were formed. Identify any stage of the process for which the study reporting is unclear (and may warrant an author query). Please draft any author questions you have on the *Author Query & Response Tab*.
- Examples:
  - » *Expected family contribution is the forcing variable used to form the intervention (Florida Student Access Grant eligibility) and comparison (Florida Student Access Grant ineligibility) groups. During the 2000/01 school year, students whose expected family contribution was less than \$1,590 were eligible for a \$1,300 Florida Student Access Grant (2000 constant dollars); this translates to families with incomes roughly below \$30,000 that year (\$40,300 in 2011 dollars) being eligible for a Florida Student Access Grant.*
  - » *The school level of free or reduced-price lunch enrollment (based on official school records) is the forcing variable used to form the intervention (teacher bonus eligibility) and comparison (bonus ineligibility) groups. Teachers in participating high-poverty schools are eligible to receive the bonus if they obtain National Board Certification. Teachers in schools with free or reduced-price lunch enrollments less than approximately the 75 percent percentile are not eligible for the bonus, whereas teachers in schools at or above the 75th percentile are eligible to receive the Challenging Schools Bonus. The Challenging Schools Bonus is an annual bonus of \$5,000.*
- **Is the study free of factors that are confounded with either group? (row 29)**
  - Select “No” in cell B29 and provide an explanation in cell C29 if there is a confounding factor; select “Yes” otherwise.
  - There is a confound if there is an  $n = 1$  problem (that is, only one unit assigned in the intervention group, the comparison group, or both), or the study suffers from another confound that would lead you to expect differences in outcomes between the groups, even if the intervention group had *not* received the intervention. Please draft any author questions you have on the *Author Query & Response Tab*.
  - Example: *(No) The study is rated Does Not Meet WWC RDD Standards. There is an  $n = 1$  confound at the interventionist level. A single teacher taught the three classes in the intervention condition. A single teacher taught the two classes in the comparison condition.*

### **Complete columns A-L of the data tab (row 30)**

Click on cell A30 to be taken to the beginning of the *Data Tab*. Complete columns M-L (instructions begin on [page 26](#)).

- **Is there at least one relevant outcome that meets the review requirements (row 31)?**
  - Select “Yes” or “No” in cell B31. List all eligible outcomes with a brief description of the measure in cell C1. Please draft any author questions you have on the *Author Query & Response Tab*.

- Examples:
  - » *(Yes) The study addressed four outcomes that meet review requirements: immediate enrollment in college assuming on-time high school graduation (that is, fall 2000); continuous enrollment into spring 2001; credits accumulated by spring 2003, three years after on-time high school graduation; and receipt of a bachelor’s degree by spring 2006, six years after on-time high school graduation.*
  - » *(No) The authors did not respond to the query for reliability information on the Trends in International Mathematics and Science Study outcome. This outcome does not meet WWC standards because it is not an established subscale, and no reliability information is provided.*

### **Standard 1: Integrity of the forcing variable (row 32)**

- **Criterion A (institutional check) (row 33).**
  - Does the description of the forcing variable and intervention assignment mechanism show that manipulation was unlikely because scorers had little opportunity or little incentive to change “true” obtained scores to allow or deny specific individuals access to the intervention? Please explain whether there was a clear *opportunity* to manipulate scores or a clear *incentive* to manipulate scores?
  - If there is no evidence of both opportunity and incentive, select “Yes” in cell B33. If there is evidence of both opportunity and incentive, select “No” in cell B33.
  - In cell C33, provide any supporting details from the study that explain the answer provided in column B. If the rating varies by outcome or sample, please specify in cell C33. Please draft any author questions you have on the *Author Query & Response Tab*.
  - Example: *(Yes) The Florida Department of Education annually sets the threshold/cutoff used to determine Florida Student Access Grant eligibility and determines an individual student’s eligibility based on the information reported in the Free Application for Federal Student Aid. The authors note that the specific expected family contribution cutoff used to determine Florida Student Access Grant eligibility was not widely or publicly advertised on either the Florida Department of Education Office of Student Financial Assistance website or individual university financial websites. The study authors completed an “exhaustive search” and found only one document from the Florida Postsecondary Planning Commission (2001) that made tangential reference to the expected family contribution cutoff used to determine Florida Student Access Grant eligibility. The study authors conclude that it is highly unlikely that students in the analytic sample would have had the opportunity or incentive to manipulate their expected family contribution, particularly given that doing so would have required considerable effort to deduce the true algorithm used to calculate the expected family contribution cutoff.*
- **Criterion B (statistical check) (row 34).**
  - Is the integrity of the forcing variable demonstrated using a statistical test found in the literature? Does that test fail to reject the null hypothesis of continuity in the density of the forcing variable at the 5 percent significance level?
  - The standards require that authors establish the smoothness of the density of the forcing variable at the cutoff using a statistical test, such as McCrary test (2008), at the 5 percent significance level.
  - In cell B34, select “Yes” if the authors have established the integrity of the forcing variable with a statistical test. Select “No” if the authors have not established the integrity of the forcing variable

with a statistical test. In cell C34, provide any supporting details. Please draft any author questions you have on the *Author Query & Response Tab*.

- Examples:
  - » *(No) The authors conducted a McCrary test to assess for smoothness of the density of the forcing variable around the cutoff. The authors did not report an exact p-value or conclusion from the hypothesis test; an author query is needed to request the actual results from the McCrary test.*
  - » *(No) The authors conducted a McCrary test to assess for smoothness of the density of the forcing variable around the cutoff. The test was statistically significant at the 5 percent significance level for the true forcing variable.*
- **Criterion C (graphical check) (row 35).**
  - Is the integrity of the forcing variable demonstrated using a graphical analysis? Is it true that there is no strong evidence of a discontinuity at the cutoff that is obviously larger than discontinuities in the density at other points?
  - The standards require that the authors present graphical evidence of the smoothness of the forcing variable outside of the expected discontinuity.
  - In cell B35, select “Yes” if the authors have illustrated the integrity of the forcing variable using graphical analysis. Select “No” if they have not. Specify figures or other supporting details in cell C35. Please draft any author questions you have on the *Author Query & Response Tab*.
  - Examples:
    - » *(Yes) The integrity of the forcing variable is demonstrated graphically in Figure 2, which presents a scatterplot of the density of observations around the cutoff. The graph clearly demonstrates the smoothness of the density around the cutoff and provides no strong evidence of other discontinuities away from the cutoff.*
    - » *(Yes) The integrity of the forcing variable is demonstrated graphically in Figure 2, which presents a density plot of observations around the cutoff. Although the graphs demonstrate some modest discontinuities in the forcing variable right around the cutoff, none of those discontinuities are obviously larger than discontinuities in the density at other points.*
- **Does the study satisfy standard 1 (integrity of the forcing variable) (row 36)?**
  - There are three possible dispositions for standard 1 (cell B36):
    - » *Completely Satisfies:* If a study satisfies all three criteria, it completely satisfies standard 1.
    - » *Partially Satisfies:* If a study satisfies two of the three criteria, it partially satisfies standard 1.
    - » *Does Not Satisfy:* If a study does not satisfy at least two of the three criteria, it does not satisfy standard 1.
  - Describe the combination of criteria the study satisfies or fails to satisfy to receive the disposition in cell C36. If the disposition varies by outcome or sample, note this in cell C36.
  - Examples:

- » *The study satisfies criteria A and C but not criterion B for standard 1; thus, the study partially satisfies standard 1.*
- » *The study satisfies criteria A, B, and C for standard 1; thus, the study completely satisfies standard 1.*

## **Standard 2: Attrition**

Complete columns M-AG of the *Data Tab* (row 38) (instructions begin on [page 28](#)). Select the appropriate attrition boundary (cautious or optimistic) in cell B39.

- **Is there at least one outcome, sample, or time point with low attrition at the cluster and subcluster levels (row 40)?**
  - On the *Main Tab* (row 40), select “No” in cell B40 if either cluster-level or subcluster-level attrition is high. If the study is an RDD or cluster RDD with low attrition, select “Yes.”
  - In cell C40, briefly describe the units of assignment and analysis and the associated attrition as necessary. In particular, provide details on how you identified the numbers used in the *Data Tab*, if this information is not readily obtained from the page numbers listed in column D. If attrition varies for different samples/analyses, briefly describe that as well. Please draft any author questions you have on the *Author Query & Response Tab*.
  - Examples:
    - » *(Yes) All outcomes, samples, and time points have low attrition at the student level, under the optimistic threshold. The authors provided baseline-assigned sample sizes for the analytic sample within the bandwidth, which permits estimation of attrition for the sample inside the bandwidth used for the impact analysis. Within the \$1,000 bandwidth used in all impact analyses, the overall attrition rate was 12.7 percent, and the differential attrition rate was 0.8 percent; thus, the study has low attrition.*
    - » *(No) Attrition is not estimable from the information reported in the study. An author query would be needed to request assigned and observed sample sizes by condition for each analytic sample of interest.*
- **Criterion A (most accurate adjustment for the forcing variable) (row 41).**
  - Is at least one of the following true?
    - » Study authors report the predicted mean attrition rate at the cutoff estimated using data from below the cutoff and the predicted mean attrition rate at the cutoff estimated using data from above the cutoff; both numbers are estimated using a statistical model that controls for the forcing variable using the same approach that was used to estimate the impact on the outcome.
    - » It is possible to calculate overall and differential attrition for the sample inside the bandwidth, with or without adjusting for the forcing variable.
  - In cell B41, select “Yes” if the authors have reported attrition in at least one of the two ways outlined in Criterion A. Select “No” if the authors have reported neither form of attrition. If attrition was calculated, briefly describe how in cell C41. Please draft any author questions you have on the *Author Query & Response Tab*.
  - Example: *(Yes) The study authors responded to an author query and provided assigned sample sizes that permitted the WWC to calculate overall and differential attrition for the sample inside the bandwidth.*

- **Criterion B (less accurate adjustment for the forcing variable)** (row 42).
  - Is it possible to calculate overall and differential attrition for the entire research sample, with or without adjusting for the forcing variable?
  - In cell B42, select “Yes” if the authors have reported enough information to calculate overall and differential attrition for the entire research sample, with or without adjusting for the forcing variable. Select “No” if the authors failed to report enough information to calculate attrition in the entire research sample. Please draft any author questions you have on the *Author Query & Response Tab*.
- **Does the study satisfy standard 2 (Attrition)?**
  - There are three possible dispositions for Standard 2 (cell B43):
    - » *Completely Satisfies*: If the study satisfies criterion A, regardless of whether it satisfies criterion B, it completely satisfies standard 2.
    - » *Partially Satisfies*: If the study satisfies criterion B but does not satisfy criterion A, it partially satisfies standard 2.
    - » *Does Not Satisfy*: If the study does not satisfy either criteria, it does not satisfy standard 2.
  - Describe the combination of criteria the study satisfies or fails to satisfy to receive the disposition in cell C43. If the disposition varies by outcome or sample, note this in cell C43.
  - Examples:
    - » *The study satisfies criterion A for standard 2; thus, the study completely satisfies standard 2.*
    - » *The study satisfies criterion B for standard 1 but does not satisfy criterion A. Thus, the study partially satisfies standard 2.*

### **Standard 3: Continuity of the relationship between the outcome and the forcing variable**

Complete columns AH-AX of the *Data Tab* (row 45). Click on cell A45 to be taken to the baseline equivalence section of the *Data Tab*. For studies that provide baseline data for the analytic sample, complete the described section (instructions begin on [page 30](#)).

- **Criterion A (baseline equivalence)** (row 46).
  - Are all of the following true?
    - » Baseline equivalence is demonstrated for at least one analytic sample by calculating an RDD impact at the cutoff on the characteristics relevant to equating the groups as given in the protocol.
    - » This analysis used either (a) exactly the same bandwidth and/or functional form as was used to estimate the impact on the outcome or (b) the same algorithm for selecting the bandwidth and/or functional form as used to estimate the impact on the outcome.
    - » The statistical model used to estimate the impact on the outcome included statistical adjustment for characteristics relevant to equating the groups as given in the protocol, if needed.
  - The list of measures on which the groups are required to be equivalent is specified in the protocol. In cell B46, select “Yes” if the study established baseline equivalence for all eligible outcomes. Select “No” if the study did not establish baseline equivalence for any of the eligible outcomes.

- In cell C46, list the outcomes that pass the baseline equivalence standard and then list the outcomes that do not pass the baseline equivalence standard. If appropriate, describe which of (a) or (b) from second option of “Are all of the following true?” the study satisfied. Please draft any author questions you have on the *Author Query & Response Tab*.
- Examples:
  - » *(Yes) As specified in the review protocol, when direct pretest or close proxies for outcomes are not available, baseline equivalence must be established for a continuous measure of baseline academic achievement and a measure of student socioeconomic status. Because none of the four outcome measures in this study have pretest values or close proxies, we therefore estimated baseline equivalence for parents’ adjusted gross income and students’ high school senior year grade point average. Baseline equivalence is satisfied for parents’ adjusted gross income (g = 0.00). Baseline equivalence also is satisfied for high school grade point average (g = 0.01). Baseline equivalence calculations were all estimated using exactly the same bandwidth (\$1,000) that was used in the impact analyses.*
  - » *(No) The authors established baseline equivalence on prior math and reading achievement test scores using the same algorithm for selecting the bandwidth (but note that the bandwidths themselves do not necessarily match). However, it is not possible to determine whether the baseline equivalence estimates are for the actual analytic sample used in the impact analyses, given that the authors do not clearly report the sample size for the analytic sample. As specified in the review protocol, for teacher retention outcomes, baseline equivalence must be established for the following measures: years of teaching experience, student achievement, and student race/ethnicity or another measure of disadvantage. The authors do not present baseline equivalence for those measures, so an author query may be needed to request the information necessary to estimate baseline equivalence for the analytic sample.*
- **Criterion B (graphical evidence) (row 47).**
  - Is there graphical evidence showing no unexplained discontinuities away from the cutoff that are larger than two times the standard error of the impact estimated at the cutoff?
  - If the authors provide graphical evidence with no large, unexplained discontinuities away from the cutoff, select “Yes” in cell B47. Otherwise, select “No” in cell B47. Provide any supporting details, such as the Figure number(s), in cell C47.
  - Examples:
    - » *(Yes) Figure 2 provides graphical analysis of the continuity in the outcome-forcing variable relationship across the values of the forcing variable. That graph provides no evidence of any discontinuities away from the cutoff.*
    - » *(Yes) Figures 3 and 4 provide graphical analysis of the continuity in the outcome-forcing variable relationship across the values of the forcing variable and the teacher retention outcome (Figure 3) and the math and reading achievement test score outcomes (Figure 4). Those graphs provide no evidence of any discontinuities away from the cutoff.*
- **Criterion C (statistical evidence) (row 48).**
  - Are all of the following true?
    - » There is statistical evidence regarding whether there are discontinuities away from the cutoff.

- » The statistical tests used the same algorithm for selecting the bandwidth/functional form as used to estimate the impact on the outcome.
- » These tests were conducted for at least four values of the forcing variable below and four values above the cutoff.
- » At least 95 percent of the impacts on the outcome at values other than the cutoff were statistically insignificant at the 5 percent significance level.
- If a sufficient number of statistical tests at the 5 percent significance level are not statistically significant, select “Yes” in cell B48. Otherwise, select “No” in cell B48. Provide any supporting details in cell C48.
- Example: *(No) The authors conducted statistical tests using the same algorithm for selecting the bandwidth used to estimate the impact on the outcomes, and the results from those analyses are reported in Table 8. Although all the statistical tests were not statistically significant at the 5 percent significance level, the authors conducted those tests for only three values (one below the actual cutoff and two above the cutoff). Thus, the study does not report sufficient information to satisfy Criterion C.*
- **Does the study satisfy standard 3 (continuity of the relationship between the outcome and the forcing variable)?**
  - There are three possible dispositions for standard 3 (cell B49):
    - » *Completely Satisfies*: If the study satisfies all three criteria, it completely satisfies standard 3.
    - » *Partially Satisfies*: If the study satisfies criterion A and either criterion B or C, it partially satisfies standard 3.
    - » *Does Not Satisfy*: If the study fails to satisfy criterion A or at least criterion B or C, it does not satisfy standard 3.
  - Describe the combination of criteria the study satisfies or fails to satisfy to receive the disposition in C49. If the disposition varies by outcome or sample, note this in cell C49.
  - Examples:
    - » *The study satisfies criteria A and B but does not satisfy criterion C. Thus, the study partially satisfies standard 3.*
    - » *The study satisfies criterion B but does not satisfy criterion A or C. Thus, the study does not satisfy standard 3.*

### **Standard 4: Functional form and bandwidth**

- **Criterion A (controlling for forcing variable) (row 51).**
  - Is the local average treatment effect for the outcome estimated using a statistical model that controls for the forcing variable?
  - If the statistical model for the outcome impact controls for the forcing variable, select “Yes” in cell B51. Otherwise, select “No” in cell B51.

- **Criterion B (local regression) (row 52).**
  - Does the study use a local regression or related nonparametric approach in which impacts are estimated within a justified bandwidth?
  - If the study uses an appropriate nonparametric approach with a justified bandwidth, select “Yes” in cell B52. If the authors do not use an appropriate parametric approach or an unjustified bandwidth, select “No” in cell B52. In cell C52, briefly describe the evidence for this response.
  - Example: *(Yes) The study authors used local linear regression to estimate impacts within a justified bandwidth. The study authors used the Imbens-Kalyanaraman (2012) bandwidth, which is a justified bandwidth selection procedure.*
- **Criterion C (“best fit” regression) (row 53).**
  - Does the study use a “best fit” regression that is shown to be a better fit to the data than at least two other functional forms using a goodness-of-fit measure from the methodological literature?
  - If the study does not use a local regression or related nonparametric approach or uses such an approach but not within a justified bandwidth, then it may estimate impacts using a “best fit” regression. If the study uses such a regression and shows sufficient evidence that it fits better than other functional forms, select “Yes” in cell B53. If the study uses a “best fit” approach but does not show it is a better fit than at least two other functional forms, select “No” in cell B53. If the study does not use a “best fit” approach, select “NA” in cell B53. In cell C53, briefly describe evidence for this response.
  - Example: *(NA) Criterion C is not applicable because the authors used local linear regression for all impact analyses.*
- **Criterion D (robustness) (row 54).**
  - Does the study provide evidence that the findings are robust to varying bandwidth or functional form choices, using at least one of the five types of evidence that are sufficient to meet this criterion?
  - If criterion B applies, select “Yes” in cell B54 under one of two conditions: The sign and significance of impact estimates are the same for at least two different justified bandwidths (such as bandwidths selected using cross-validation and the Imbens and Kalyanaraman [2012] method), or the sign and significance of impact estimates are the same for at least one justified bandwidth and at least two additional unjustified bandwidths.
  - If criterion C applies, select “Yes” in cell B54 under one of two conditions: The sign and significance of the impact estimates are the same using two different goodness-of-fit measures to select functional form, or the sign and significance of the impact estimates must be the same for at least three different functional forms (including the “best fit” form).
  - If both criteria B and C apply, select “Yes” in cell B54 if the sign and significance of impact estimates are the same for the impact estimated within a justified bandwidth and the impact using a “best fit” regression. Otherwise, select “No” in cell B54. In cell C54, briefly describe evidence for this response.
  - Example: *(Yes) The study provides evidence that findings for the receipt of bachelor’s degree outcome are robust to varying bandwidth and functional form choices (with the results from those sensitivity analyses shown in Table 7). The sign (positive) and significance (statistically significant) of the impact estimate for*

*this outcome are the same for a total of four different bandwidths and one different functional form. The authors do not report findings from those sensitivity analyses for the other outcomes in the study, but they state in the footnote on page 27 that the results were robust to those different specifications.*

- **Criterion E (graphical analysis) (row 55).**
  - Is there a graphical analysis displaying the relationship between the outcome and forcing variable, including a scatterplot and a fitted curve, and is the display not obviously inconsistent with the choice of bandwidth and functional form used to estimate the impact on the outcome?
  - If the study authors provide an appropriate graphical display of the relationship between the outcome and the forcing variable, select “Yes” in cell B55. Otherwise, select “No” in cell B55. In cell C55, briefly describe evidence for this response.
  - *Example: (No) The study authors do not provide a scatterplot with “best fit” curve(s) to displaying the relationship between the outcomes and the expected family contribution forcing variable. Although Figure 3 provides a scatterplot displaying predicted values within bins, the scatterplot does not display the raw data. Thus, Figure 3 does not suffice to satisfy this criterion.*
- **Criterion F (constraints) (row 56).**
  - Is it true that the relationship between the forcing variable and the outcome is not constrained to be the same on both sides of the cutoff?
  - If the statistical model(s) used in the study allow the relationship between the forcing variable and the cutoff to vary on either side of the cutoff, select “Yes” in cell B56. Otherwise, select “No” in cell B56.
- **Does the study satisfy standard 4 (functional form and bandwidth)?**
  - There are three possible dispositions for standard 4 (cell B57):
    - » *Completely Satisfies:* If the study satisfies criteria A, B, D, E, and F, then this study completely satisfies standard 4.
    - » *Partially Satisfies:* If this study satisfies criterion A, either of criterion B or C, and criterion E, then this study partially satisfies standard 4.
    - » *Does Not Satisfy:* In all other cases, this study does not satisfy standard 4.
  - Describe the combination of criteria the study satisfies or fails to satisfy to receive the disposition in cell C57. If the disposition varies by outcome or sample, note this in cell C57.
  - Examples:
    - » *The study does not satisfy criterion D but satisfies criteria A, B, C, E, and F. Thus, the study partially satisfies standard 4.*
    - » *The study satisfies criteria A-D and F for the degree receipt outcome. However, the study does not satisfy criterion E. Thus, the study does not satisfy standard 4.*

## Standard 5: Fuzzy regression discontinuity design (FRDD)

In a sharp RDD, all intervention group members receive intervention services and no comparison group members receive intervention services. In an FRDD, some intervention group members do not receive intervention services and/or some comparison group members do receive intervention services, but there is still a substantial discontinuity in the probability of receiving services at the cutoff.

Eight criteria determine whether a study satisfies this standard. All eight criteria are waived for impact estimates calculated using a reduced form model (in which the outcome is modeled as a function of the forcing variable, an indicator for being above or below the cutoff, and possibly other covariates, but the *participation* indicator is not included in the model).

- If the RDD is sharp, or if the FRDD impact estimates are calculated using a reduced form model, select “Waived” in cells B59–B66.
- **Criterion A** (row 59).
  - Is the participation indicator a binary indicator for taking up at least a portion of the intervention?
  - If the participation indicator is binary and indicates taking up at least a portion of the intervention, select “Yes” in cell B59. Otherwise, select “No” in cell B59.
- **Criterion B** (row 60).
  - Does the estimation model have exactly one participation indicator?
  - If the estimation model has exactly one participation indicator, select “Yes” in cell B60. Otherwise, select “No” in cell B60.
- **Criterion C** (row 61).
  - Is the indicator for being above or below the cutoff a binary indicator for the groups to which study participants are assigned?
  - If indicator is binary, select “Yes” in cell B61. Otherwise, select “No” in cell B61.
- **Criterion D** (row 62).
  - Are the same covariates included in the analysis that estimates the impact on participation and the analysis that estimates the impact on the outcome?
  - If the same covariates (including the forcing variable) are included in the analysis that estimates impacts on participation and the analysis that estimates impacts on outcomes, select “Yes” in cell B62. Otherwise, select “No” in cell B62.
- **Criterion E** (row 63).
  - Are there no clear violations of the exclusion restriction?
  - To satisfy this requirement, intervention assignment must have no impact on the outcome except in the take-up of the intervention. In addition, participation between the assigned intervention and the assigned comparison group must be defined consistently. If a study satisfies the exclusion restriction, select “Yes” in cell B63. Otherwise, select “No” in cell B63. In cell C63, briefly describe the evidence for this response.

- **Criterion F (row 64).**
  - Is the forcing variable shown to be a strong predictor of participation in the intervention; that is, does it meet the minimum required first-stage  $F$ -statistic or  $t$ -statistic?
  - In a regression of program participation on an intervention indicator and other covariates, if the intervention indicator reports a minimum  $F$ -statistic of 16 or a minimum  $t$ -statistic of 4, select “Yes” in cell B64. For FRDD studies with more than one indicator for being above or below the cutoff, see the WWC RDD standards for randomized controlled trials that report complier average causal effect estimates for the minimum required first-stage  $F$ -statistic. Otherwise, select “No” in cell B64.
- **Criterion G (row 65).**
  - Does the study use a local regression or related nonparametric approach in which FRDD impacts are estimated within a justified bandwidth and is at least one of the following true?
    - » The bandwidth is justified for the FRDD impact estimate (that is, the ratio).
    - » The study uses separate justified bandwidths for the numerator and the denominator.
    - » A justified numerator bandwidth is used for the FRDD impact estimate and is less than or equal to a justified bandwidth for the denominator.
  - If the study has justified bandwidth for the entire FRDD estimate, has a separate justified bandwidth for both the numerator and the denominator, or uses a bandwidth for both the numerator and the denominator that is justified for the numerator and is no wider than a justified bandwidth of the denominator, then select “Yes” in cell B65. In cell C65, briefly describe the evidence for this response, including which of the three possible subcriteria the study meets.
- **Criterion H (row 66).**
  - Is at least one of the following true?
    - » The FRDD impact is estimated using a bandwidth that is justified only for the numerator.
    - » The denominator is estimated using a “best fit” functional form that is shown to be a better fit to the data than at least two other functional forms using a goodness-of-fit measure from the methodological literature.
  - If the FRDD impact is estimated using a bandwidth that is justified only for the numerator or the denominator is estimated using a “best fit” functional form with sufficient evidence of its fit, select “Yes” in cell B66. Otherwise, select “No” in cell B66. In cell C66, briefly describe the evidence for this response, including which of the two possible subcriteria the study meets.
- **Does the study satisfy standard 5 (FRDD)? (row 67).**
  - There are three possible dispositions for standard 5 (cell B67)
    - » *Completely Satisfies*: If all of the criteria are waived because the study is a sharp RDD or a FRDD using a reduced form, this study completely satisfies standard 5. If the study satisfies criteria A-G, this study completely satisfies standard 5.
    - » *Partially Satisfies*: If this study satisfies criteria A-F and satisfies criterion H, then this study partially satisfies standard 5.

- » *Does Not Satisfy*: If this study fails to meet any of criteria A-G and/or does not meet either of criterion G or H, this study does not satisfy standard 4.
- Describe the combination of criteria the study satisfies or fails to satisfy to receive the disposition in cell C67. If the disposition varies by outcome or sample, note this in cell C67.
- Examples:
  - » *The study authors used a sharp RDD, so the study completely satisfies standard 5.*
  - » *FRDD criteria are waived given that the authors estimated impact estimates using a reduced form model.*
- **Is the study free of other data or analytic issues that would affect the rating?** (row 68).
  - In cell B68, select “No” if there are issues about the analysis or data that were not captured in an earlier entry. Use column C to summarize the issues. If there are no issues, select “Yes.”
  - Example: *The degrees of freedom reported for the F-statistics are not consistent, suggesting that there is variation in the number of students providing each outcome. For the most part, they appear to be within 10 students of the sample sizes reported in the article and used in the Data Tab to assess attrition. Rows 30-34 were used to determine whether the attrition rating would change if the 10 students were removed equally from both groups (row 30), from intervention (row 31), or from comparison (row 32); if three fifths from intervention and two fifths from comparison were removed (row 33), or if two fifths from intervention and three fifths from comparison were removed (row 34). Regardless of the distribution of the sample loss, the study still has low attrition of youth.*
- **What is the highest rating of any main finding in the study given current information?** (row 69).
  - Select a rating using the drop-down menu in cell B69.
  - If you select *Meets RDD Standards With (or Without) Reservations*, in cell C69 select “The study meets WWC Regression Discontinuity Design Standards with or without reservations. An explanation for the highest rating is below.” In row 70, briefly describe the contrast(s) and rating(s). You also will need to complete stage 3.
  - Example: *(Does Not Meet RDD Standards) The study is rated Does Not Meet WWC Regression Discontinuity Design Standards for the contrasts between Challenging Schools Bonus-eligible and Challenging Schools Bonus-ineligible schools on the following three outcomes: teacher retention, student math achievement, and student reading achievement. The study received this rating because the study completely satisfies RDD standard 1, partially satisfies standard 4, and does not satisfy standards 2 or 3. Insufficient information is provided in the report that permits assessment of the sample attrition and continuity standards; an author query should solicit that information (which, if provided, could potentially change the study’s disposition).*
  - If you select *Does Not Meet RDD Standards*, do not complete stage 3.
  - If the study is rated *Does Not Meet RDD Standards* based on the information provided, select the appropriate Disposition Code in cell C69 and provide an explanation for the disposition code in row 70. Copy and paste row 69 as many times as needed to capture the exact disposition code (selected from menu in cell C69) for each comparison that could be rated *Does Not Meet WWC RDD Standards*.

- The *Does Not Meet WWC RDD Standards* Disposition Codes are as follows:
  - » ... **the measures of effectiveness cannot be attributed solely to the intervention.** This covers scenarios in which a study had only one unit assigned to one or both conditions, or the effects of the intervention of interest were reported only in combination with other interventions.
  - » ... **the eligible outcomes do not meet WWC requirements.** This covers scenarios in which the outcomes were overaligned with the intervention, the outcomes were determined not to be sufficiently valid or reliable, or interassessor agreement did not meet minimum thresholds.
  - » ... **the integrity of the forcing variable is not established.** This covers scenarios in which there is reason to believe that there may have been both opportunity and incentive to manipulate values of the forcing variable, or the integrity of the forcing variable was not established statistically or graphically.
  - » ... there is a high or unknown level of attrition, and the study does not demonstrate continuity of the outcome-forcing variable relationship. This covers scenarios in which the study provided neither of the two acceptable form attrition calculation and failed to establish baseline equivalence on key covariates, or the continuity of the relationship between the outcome and forcing variable was not established graphically or statistically.
  - » ... **there is no control for the forcing variable in the statistical model.** This covers scenarios in which the impact model did not include the forcing variable as a covariate.
  - » ... **the study uses neither local regression within a justified bandwidth nor a “best fit” regression.** This covers scenarios in which the study did not use a justified bandwidth or did not provide sufficient evidence that the functional form was the “best fit” for the data.
  - » ... there is no graphical analysis displaying the relationship between the outcome and forcing variable that demonstrates the model fits the data well. This covers scenarios in which there is no graphical display of the relationship of the raw outcome and forcing variable data that demonstrates the relationship between the outcome and the forcing variable.
  - » ... the participation indicator is not a binary indicator for taking up at least a portion of the intervention. This covers FRDD scenarios in which the participation indicator is not binary.
  - » ... **the estimation model does not have exactly one participation indicator.** This covers FRDD scenarios in which participation indicators indicate multiple degrees of participation.
  - » ... the indicator for being above or below the cutoff is not a binary indicator for the groups to which study participants are assigned. This covers FRDD scenarios in which the indicator for groups is not binary—for instance, if there were multiple indicators for varying degrees of participating in treatment.
  - » ... the same covariates are not included in the analyses that estimate the impact on participation and the impact on the outcome. This covers FRDD scenarios in which the participation and outcome models differ in their included covariates.
  - » ... **there is a clear violation of the exclusion restriction.** This covers FRDD scenarios in which participation is defined differently between the assigned intervention and assigned intervention

groups, or being assigned has impacts on the outcome other than through the take-up of the intervention.

- » ... **the forcing variable is not shown to be a strong predictor of participation in the intervention.** This covers FRDD scenarios in which the strength of the relationship between the forcing variable and participation in the intervention either is not provided or is insufficiently strong.
- » ... **the FRDD impact estimate does not use a justified bandwidth or the denominator does not use a “best fit” functional form.** This covers scenarios in which the total FRDD estimate bandwidth is unjustified; the bandwidth for the denominator is not justified; or a justified bandwidth for the numerator is used for the FRDD estimate that is wider than the justified bandwidth for the denominator, and the denominator is estimated using a “best fit” functional form with insufficient evidence.
- » ... **the analysis does not provide a credible measure of the effectiveness of the intervention.** This covers scenarios in which other analytical issues may affect the rating, including scenarios such as a cluster RDD in which individual-level SDs are not provided to establish baseline equivalence.
- If an author query is needed to determine the final rating, indicate what response would be needed. The author query should be referenced in the appropriate row above and the question drafted on the *Author Query & Response Tab*.
- If the rating differs by analysis, provide the rating for each sample, outcome, and time period combination and describe the differences in row 72.
- Examples:
  - » *The study is rated Does Not Meet WWC Regression Discontinuity Design Standards for the contrasts between Florida Student Access Grant-eligible and Florida Student Access Grant-ineligible students on the following four outcomes: immediate college enrollment, continuous college enrollment, college course credit accumulation, and college completion. The study received this rating because the study completely satisfies RDD standards 2 and 5; does not satisfy RDD standard 4; and partially satisfies RDD standards 1 and 3.*
  - » *The study is rated Does Not Meet WWC Regression Discontinuity Design Standards for the contrasts between Challenging Schools Bonus-eligible and Challenging Schools Bonus-ineligible schools on the following three outcomes: teacher retention, student math achievement, and student reading achievement. The study received this rating because the study completely satisfies RDD standard 1, partially satisfies standard 4, and does not satisfy standards 2 or 3. Insufficient information is provided in the report that permits assessment of the sample attrition and continuity standards; an author query should solicit that information (which, if provided, could potentially change the study’s disposition).*

### Stage 3: Study details

Stage 3 of the SRG summarizes the key findings and a broad description of the study design and intervention. 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 individual study review. As such, do not use any text that was directly copied from the study in the descriptions of the study details in stage 3 and ensure that the text is written in IES style.

Complete columns AY-BQ of the *Data Tab* (row 75): Click on cell A75 to be taken to the study-reported and WWC-computed findings sections of the *Data Tab* (instructions begin on [page 32](#)).

- **Did the authors present effect sizes?** If so, how were they computed? (row 76).
  - If the study authors present effect size estimates, select “Yes” in cell B76 and indicate how those effect sizes were computed in cell C76 of the *Main Tab* (regardless of whether impact estimates are provided in other metrics).
  - If the authors do not report effect sizes, select “No.”
  - In particular, note whether the author-reported effect size differs from the WWC-calculated effect size and provide information that explains why (for example, the authors use the comparison group SD, not the pooled SD, in their calculations).
  - Examples:
    - » *The authors do not report effect sizes.*
    - » *The authors report effect sizes; however, they use the comparison group SD in the denominator rather than the full sample SD. This results in slight differences in magnitude from the WWC-calculated effect sizes.*
- **Are estimates presented for subgroups in protocol?** (row 77).
  - In cell B77, select “Yes” if the study provides impact estimates for any of the subgroups outlined in the protocol under *Eligible Populations*; select “No” otherwise.
  - In cell C77, briefly describe which subgroups are analyzed and whether those analyses meet WWC Regression Discontinuity Design Standards. In the *Data Tab*, be sure to include information from subgroup analyses in addition to the main analyses. (See the instructions on completing the *Data Tab* for more information.)
- **Contextual variables** (row 78). Select the appropriate entry from the dropdown box or enter a value in column B for each contextual variable. Select “not reported” in the dropdown box or record “NR” for any contextual variables not reported in the study. Percentages should be entered up to two decimal points; any entries with more than two decimal points will be automatically rounded. Briefly describe any important context in column C.
  - **Number of students** (row 79): Enter the number of students within the analytic sample in cell B79. If the number of students in the analytic sample varies by subgroup or outcome, provide details in cell C79.
  - **Grades** (row 80): Select one option from the drop-down menu in cell B80: *Pre-K, Kindergarten, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, Postsecondary, Adult, Mixed grades, or Not reported*. If the *Mixed grades* option is selected, list additional details about the grades covered in cell C84.
  - Percentage White; Black; Native American/Alaska Native; Asian; Pacific Islander; two or more races; and other or unknown race (rows 81-87).
  - Percentage Hispanic; Non-Hispanic; and other or unknown ethnicity (rows 88-90).
  - Percentage male; female; and other or unknown gender (rows 91-93).

- Percentage English learner students; not English learner students; and other or unknown English learner status (rows 94-96).
- Percentage individualized education program (IEP)/special education; not individualized education program (IEP)/special education; and other or unknown individualized education program (IEP)/special education (rows 97-99).
- Percentage 504 plan; not 504 plan; and other or unknown 504 plan (rows 100-102).
- Percentage autism spectrum disorder; deaf-blindness; developmental delay; emotional disturbance; hearing impairment; intellectual disability; multiple disabilities; orthopedic impairment; other health impairment; specific learning disability; speech or language impairment; traumatic brain injury; visual impairment; no disabilities; and other disability (rows 103-117).
- Percentage free or reduced-price lunch; no free or reduced-price lunch; and other or unknown free or reduced-price lunch (rows 118-120).
- Percentage Pell Grant eligibility; no Pell Grant eligibility; and other or unknown Pell Grant eligibility (rows 121-123).
- **Type of classroom** (row 124): Select one option from the drop-down menu in cell B124: *General education, Self-contained special education classroom, Both general education and self-contained special education classrooms, or Not reported.*
- **Type of school** (row 125): Select one option from the drop-down menu in cell B125: *Public, Private, Parochial, Charter, Mixed school type, or Not reported.* If more than one type of school was included, list all types in cell C125.
- **Urbanicity** (row 126): Select one option from the drop-down menu in cell B126: *Rural, Suburban, Town, Urban, Mixed urbanicity, or Not reported.* If more than one geographical setting was included in the study, list additional settings in cell C126. Choices for urbanicity in the SRG are based on categories in the urban-centric classification system developed by the National Center for Education Statistics (see <https://nces.ed.gov/surveys/ruraled/definitions.asp>):
  - » *Rural*: Census-defined rural territory that does not lie inside an urbanized area or urban cluster.
  - » *Suburban*: Territory inside an urbanized area (with population of 50,000 or more) but outside a principal city.
  - » *Town*: Territory inside an urban cluster (with populations between 2,500 to 50,000).
  - » *Urban*: Territory inside an urbanized area (with population of 50,000 or more) and inside a principal city.
  - » If the study involved more than one type of urbanicity, list all urbanicities in cell C126.
- **Location** (row 127): Select one option from the drop-down menu in cell B127: *West region, Northeast region, Midwest region, South region, an individual listing of all 50 U.S. states, individual listings of U.S. territories and affiliated states, Mixed locations, or Not reported.* If the study took place in more than one region, list all regions in cell C127.

- **Delivery method** (row 128): Select one option from the drop-down menu in cell B128: *Individual, Small group, Whole class, School, Mixed delivery method, or Not reported*. If more than one delivery method was used, note it in cell C128.
- **Intervention type** (row 129): Select one option from the drop-down menu in cell B81: *Practice, Product, Policy, Program, Mixed intervention type, or Not reported*. If more than one intervention type was used, list all intervention types in cell C129.
- **Intervention scope** (row 130): Select one option from the drop-down menu in cell B130: *Curriculum, In-school supplemental, Out-of-school supplemental, Mixed intervention scope, or Not reported*. If more than one intervention scope was used, list all intervention types in cell B130.
- **Intervention target** (row 131): Select one option from the drop-down menu in cell B131: *Student-focused, Teacher-focused, School leader-focused, Mixed intervention target, or Not reported*. If more than one intervention target was used, list all intervention targets in cell B131.
- **Setting of the study** (for example, location, classrooms, courses, schools) (row 132): Include the locations from which the sample was drawn. Example: *The study took place in Washington state public high schools*.
- **Study design**. (row 133): Summarize the study design, including the forcing variable and how it separated participants into the assigned intervention and assigned comparison groups. Examples:
  - » *The school level of free or reduced-price lunch enrollment (based on official school records) is the forcing variable used to form the intervention (teacher bonus eligibility) and comparison (bonus ineligibility) groups. Teachers in participating high-poverty schools are eligible to receive the bonus if they obtain National Board Certification. Teachers in schools with free or reduced-price lunch enrollments less than approximately the 75th percentile are not eligible for the bonus, whereas teachers in schools at or above the 75th percentile are eligible to receive the CBS. The Challenging Schools Bonus is an annual bonus of \$5,000.*
  - » *The study uses a sharp RDD. During the 2000/01 school year, students whose expected family contribution was less than \$1,590 were eligible to receive a \$1,300 Florida Student Access Grant. The expected family contribution cutoff used to determine Florida Student Access Grant eligibility was approximately \$30,000 that year (or roughly \$40,300 in 2011 dollars). Participants in the study included high school seniors in Florida public high schools during the 1999/2000 school year who submitted a Free Application for Federal Student Aid application.*
- **Sample sizes** (for example, students, classrooms, teachers, schools) (row 134): Summarize the participants, including the characteristics of the participant sample. Include a description of the number of clusters/students in the analytic sample. Examples:
  - » *The intervention condition included 10 schools, 25 classrooms, and 1,500 students. The analytic sample included 10 schools, 25 classrooms, and 1,000 students. The comparison condition included 10 schools, 20 classrooms, and 1,350 students. The analytic sample included 9 schools, 18 classrooms, and 1,200 students. There is a single outcome, so there is no variation in analytic sample sizes by outcome.*
  - » *The sample includes public high school students and teachers in Washington. The analytic sample varies across outcome and analytic models but generally includes roughly 6,792 participants. The study does not report the number of schools these students attended.*

- **Sample characteristics in protocol (for example, race, gender, free or reduced price lunch) (row 135):** Examples: *The analytic sample for the intervention condition included 1,000 students. The analytic sample was 51 percent female, 35 percent Black, 30 percent Asian, 25 percent Caucasian, and 10 percent Hispanic. Eighty percent of the students were eligible for free or reduced-price lunch. Twenty percent were receiving special education or related services but were enrolled in general education history classes. The analytic sample for the comparison condition included 1,200 students. The student sample was 49 percent female, 35 percent Caucasian, 25 percent Black, 20 percent Asian, and 20 percent Hispanic. Seventy percent of the students were eligible for free or reduced-price lunch. Fifteen percent were receiving special education or related services but were enrolled in general education history classes.*
- **Intervention condition as implemented in the study (including the number of days/weeks/months, number of sessions, time per session) (row 136):** Summarize the intervention(s) in sufficient detail to help readers understand what makes this intervention similar to or different from other interventions. The level of detail should be similar to what would be provided in an introductory section of a typical impact evaluation report. It should include the length of the intervention and the dosage as well as information about the content, delivery, and implementation of the intervention. This description should be about the intervention as used in this study, not as described by the developer or in ideal conditions. Example: *The Challenging Schools Bonus offers an annual bonus of \$5,000 to teachers working in high-poverty schools, defined as schools with a high proportion of students eligible for free or reduced-price lunch, who earn certification through the National Board for Professional Teaching Standards.*
- **Comparison condition as implemented in the study (row 137):** Indicate the consequences of being assigned to the comparison group (for example, what comparison group members could not receive, what the study suggests they did receive). Clarify whether the counterfactual was a particular alternative intervention, and if so, name the intervention and provide a brief description, if the study provides that information. Example: *The Challenging Schools Bonus offers an annual bonus of \$5,000 (U.S. dollars) to teachers working in high-poverty schools, defined as schools with a high proportion of students eligible for free or reduced-price lunch, who earn certification through the National Board for Professional Teaching Standards. Comparison schools were below the free or reduced-price lunch eligibility threshold and thus were not eligible for the \$5,000 Challenging Schools Bonus teacher bonus.*
- **Describe all eligible outcomes reported and how they were measured (row 138).**
  - Describe all the outcomes within relevant domains examined in the study and identify which of those outcomes are eligible based on the criteria specified in the protocol under *Types of Outcomes to Be Included*. Also, indicate how each eligible outcome was measured (if it is not self-explanatory from the name of the outcome), how it was collected (if relevant), and how it can be interpreted (the scale of the measure).
  - Example: *The study reports findings for four eligible outcomes under this review protocol:*
    - » *Immediate enrollment in college, assuming on-time high school graduation (that is, fall 2000).*
    - » *Continuous enrollment into spring 2001.*
    - » *Credits accumulated by spring 2003, three years after on-time high school graduation.*

- » *Receipt of a bachelor's degree by spring 2006, six years after on-time high school graduation.*
- » *All of these outcomes were measured using official state records data from the Florida Department of Education's K-20 Data Warehouse.*
- **Are there outcomes that do not meet review requirements?** If yes, provide the domain and a brief description of the reason (row 139).
  - Select “Yes” or “No” in cell B107. In cell C107, list all eligible outcomes and why they do not meet review requirements.
  - Example: *The authors did not respond to the query for reliability information on the Trends in International Mathematics and Science Study outcome. This outcome is based on 20 publicly available items but is not an established subscale. The Trends in International Mathematics and Science Study outcome does not meet WWC standards because it is not an established subscale, and no reliability information is provided.*
- **Are any outcomes not eligible for review?** If yes, provide a brief description and the reason (row 140).
  - Select “Yes” or “No” in cell B108. In cell C108, list the outcomes that are not eligible for review, along with a reason for ineligibility.
  - Example: *The study reports findings on several outcomes there are not eligible for review because they are subsets of the overall eligible outcomes (for example, initial enrollment at public vs. private; two-year vs. four-year) and at different follow-up assessments.*
- **Support for implementation** (row 141): Indicate both the staff training and technical assistance conducted to support the implementation of the intervention (as evaluated in the study).

## **DATA TAB OF THE SRG**

Each row corresponds to a single outcome for a particular sample at a particular point in time. Pre- and postintervention outcomes using the same measure and sample should be reported in a single row. Most of the column headings have a comment attached that provides a short description of the information to be captured within it.

### **Outcome name, domain, and measure characteristics section of the data tab (columns A-L)**

#### **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 may provide on outcome measures.
- If results are presented separately for different samples or periods, you will need to have a separate row for each. Similarly, if there are multiple follow-ups, include a row for each. If reliability information differs across the samples, make sure it is accurately reflected in the table.
- Measures that are used only as a preintervention measure (such as age or gender) or postintervention measure (such as dropout) are recorded on a separate row and labeled accordingly.

## Characteristics

- **Measure** (column A): Indicate the name of the outcome or test, exactly as it was specified in the article or report.
- **Domain** (column B): Indicate the eligible domain(s) under which the outcome falls.
  - If a study reports subgroup analyses, composite and subtests, or outcomes at different points in time, then the domain name in the *Data Tab* must reflect this to ensure the domain averages and multiple comparisons are conducted correctly. In taking this approach, the main analyses will not use multiple comparisons when the authors present subgroup, subtest, or additional follow-up examples. In the presentation of the results in an intervention report or single-study report, the evidence rating for the study can be based on the main analyses and the additional analyses can be included in appendices for transparency (though these findings will not contribute to the evidence rating).
  - For example, in subgroup analyses (which are recorded in the *Data Tab* below the full sample analysis), the “Math” domain in the full sample analysis could be named the “Math-Special Education” domain for a special education subgroup analysis.
  - For studies that report outcomes for both a composite measure and multiple subtests, the SRG could use the “Math” domain for the composite and “Math-Subtests” for the arithmetic, fractions, and whole number subtests.
  - Similarly, for an outcome in the Math domain collected at three time points, the SRG could use “Math” for the time period that will contribute to the evidence and “Math-X mo” for the time periods that will be reported in appendices.
  - If an outcome or contrast is rated *Does Not Meet WWC RDD Standards* and will not be included in the report, the domain column should be empty to ensure the multiple comparison adjustments are correct.
- **Binary** (column C): 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” (that is, there is no underlying distribution of the variable). Select “No” otherwise.
- **Standardized test** (column D): 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 E): Select “Yes” from the drop-down menu if the measure appears to be a reasonable measure; select “No” if you see an obvious problem with the measure and indicate your concerns in column L.
- **Test-retest reliability** (column F): Enter the test-retest reliability of the outcome if it is reported; enter “NR” otherwise. The minimum acceptable value under the Study Review Protocol is 0.60.
- **Internal consistency** (column G): Enter the internal consistency of the outcome if it is reported; enter “NR” otherwise. The minimum acceptable value under the Study Review Protocol is 0.60.
- **Interrater reliability** (column H): Enter the interrater reliability of the outcome if it is reported; enter “NR” otherwise. The minimum acceptable value under the Study Review Protocol is 0.60.

- **Not overaligned?** (column I).
  - Select “Yes” from the drop-down menu if you have no concerns that the measure may be overaligned with the intervention.
  - Select “No” from the drop-down menu if you have concerns that the measure may be overaligned with the intervention.
  - Measures that are closely aligned or tailored to the intervention are likely to demonstrate larger effect sizes than those that are less closely aligned with the intervention. An example of overalignment is if the measure includes some of the same materials (such as specific reading passages) that are used in the intervention or administered to the intervention group as part of the intervention. Explain any concerns in column L.
- **Same measure and collection?** (column J).
  - Select “Yes” from the drop-down menu if the same measure was collected similarly for the intervention and comparison groups.
  - Select “No” from the drop-down menu if it is clear that outcome data were collected differently for the intervention and comparison groups, potentially in a way that could lead to differences in average outcomes between groups.
  - In a situation in which the outcome data were collected differently across the intervention and comparison groups, this may be considered a confounding factor and documented in the *Main Tab* (row 29). Explain any concerns in column L.
- **Meets WWC requirements?** (column K).
  - 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 column L includes details on why the measure does not meet WWC requirements.
- **Notes or concerns about the measure** (column N).
  - Summarize any concerns you have about this measure. In particular, if the measure does not meet the requirements for inclusion in this review as noted in column K, indicate why.
  - If a citation is provided for a standardized test or subscale, it may be helpful to note that here.

Now, go back and enter a decision in cell B31 on the *Main Tab*. (Is there at least one relevant outcome that meets review requirements?)

### **Attrition section of the data tab (columns M-AG)**

For each outcome, enter the necessary sample information in that row as follows.

- **Using data from...** (column M): For studies that provide predicted rates of attrition at the cutoff, select “Predicted proportions at cutoff,” or “Predicted percentages at cutoff.” These options are provided for

ease of use, and percentages also can be transformed into proportions by dividing the percentage by 100. For studies that provide sample sizes, select “Unadjusted rates.”

For cluster RDD studies, complete columns N-V.

- **Unit of assignment (if different)** (column N): For clustered designs, enter the unit of assignment (teacher, student).
- If attrition was calculated using predicted rate (percentages or proportions) at the cutoff,
  - **Predicted cluster rates at cutoff** (columns O and P).
    - » In column O, enter the predicted rate of attrition at the cutoff for all the units eligible to be assigned to the intervention group.
    - » In column P, enter the predicted rate of attrition at the cutoff for all the units eligible to be assigned to the comparison group.
- If attrition was calculated using assigned and analyzed sample sizes,
  - **Assigned sample clusters** (columns Q and R).
    - » In column Q, enter the number of clusters at the cutoff eligible to be assigned to the intervention group.
    - » In column R, enter the number of clusters at the cutoff eligible to be assigned to the intervention group.
  - **Analytic sample clusters** (columns S and T).
    - » In column S, enter the number of clusters for the intervention group in the analytic sample.
    - » In column T, enter the number of clusters for the comparison group in the analytic sample.
  - **Attrition rates** (columns U and V): The overall attrition rate at the cluster level is calculated in column U. The differential attrition (in percentage points) is calculated in column VV.
  - **Low at cluster level?** (column W): A determination of whether attrition is low at the cluster level is reported (“Yes” or “No”).

For all studies, complete columns X-AG.

- **Unit of analysis** (column X): Enter the unit of analysis (also known as the *subcluster*).
- If attrition was calculated using predicted rates (proportions or percentages) at the cutoff:
  - **Predicted rates at cutoff** (columns Y and Z).
    - » In column Y, enter the predicted rate of attrition at the cutoff for all units eligible to be assigned to the intervention group.
    - » In column Z, enter the predicted rate of attrition at the cutoff for all units eligible to be assigned to the comparison group.
- If attrition was calculated sample sizes:
  - **Assigned sample** (columns AA and AB).

- » In column AA, enter the sample size for units eligible to be assigned to the intervention group.
- » In column AB, enter the sample size units at the cutoff eligible to be assigned to the comparison group.
- **Analytic sample** (columns AC and AD).
  - » In column AC, enter the sample size at the cutoff for the intervention group for the analytic sample.
  - » In column AD, enter the sample size at the cutoff for the comparison group for the analytic sample.
- **Attrition rates** (columns AE and AF): The overall attrition rate for the unit of analysis is calculated in column AE. The differential attrition for the unit of analysis (in percentage points) is calculated in column AF.
- **Low?** (column AG): A determination of whether attrition is low for the unit of analysis is reported (“Yes” or “No”).

Now go back and enter a decision in cell B40 on the *Main Tab* (Is there at least one outcome, sample, or time point with low attrition at the cluster and subcluster level?).

### Baseline equivalence section of the data tab (columns AH–AW)

- **Using data from ...** (column AH): Select the type of data that will be used to assess baseline equivalence.
  - For continuous measures, authors must estimate the impact at the cutoff on the covariate of interest, and the study must either use exactly the same bandwidth and/or functional form as was used to estimate the impact on the outcome or use the same algorithm for selecting the bandwidth and/or functional form as was used to estimate the impact on the outcome.
  - Impact and full sample SD should be selected if the authors report an impact instead of predicted mean values for the two groups. Adjusted M and full sample SD should be selected if the authors provide predicted means at the cutoff instead of an impact estimate.
  - “Dichotomous” should be selected if the authors provide the predicted mean covariate value—that is, the predicted probability—at the cutoff estimated using data from below the cutoff and the predicted probability at the cutoff estimated using data from above the cutoff. Both predicted probabilities must be calculated using the same statistical model that is used to estimate the impact on the covariate at the cutoff.
- **Intervention** (columns AI–AL).
  - If the baseline measure was a dichotomous outcome: Enter the preintervention proportion for the intervention group in AI. This should be entered as a decimal value (that is, “0.50,” not “50” for 50 percent).
  - If the baseline measure was a continuous outcome with adjusted means and full-sample standard deviations reported for both the intervention and comparison groups:
    - » In column AJ, enter the preintervention mean for the intervention group.

- » In column AK, enter the full-sample preintervention SD for the intervention group.
- » In column AL, enter the preintervention sample size for the baseline equivalence calculation. This may be different from the values entered in the attrition calculation section, depending on how the authors calculated baseline equivalence.
- **Comparison** (columns AM-AP).
  - If the baseline measure was a dichotomous outcome: In column AM, enter the preintervention proportion for the comparison group. This should be entered as a decimal value (that is, “0.50,” not “50” for 50 percent).
  - If the baseline measure was a continuous outcome with adjusted means and full-sample SDs reported for both the intervention and comparison groups:
    - » In column AN, enter the preintervention mean for the comparison group.
    - » In column AO, enter the full-sample preintervention SD for the comparison group.
    - » In column AP, enter the preintervention sample size for the comparison group for the baseline equivalence calculation. This may be different from the values entered in the attrition calculation section, depending on how the authors calculated baseline equivalence.
- **Impact** (column AQ): If the baseline measure was a continuous outcome reported as a covariate-adjusted impact:
  - In columns AK and AO, enter the full-sample preintervention SD for the intervention and comparison groups, respectively.
  - In column AQ, enter the covariate-adjusted impact estimate.
- **g** (column AS): Based on the information entered in columns AH-AQ, the magnitude of the baseline difference will be reported as Hedges’ *g*.
- **Equivalent?** (column AT): Based on the information calculated in column AS, the determination of equivalence will be provided.
  - “Yes” indicates the magnitude of the baseline difference is less than 0.05 SD.
  - “No” indicates the magnitude of the baseline difference is greater than 0.25 SD.
  - “Adj” indicates the authors must report findings from an analysis that statistically controlled for the baseline difference.
- **Was the pretest different?** (column AU).
  - Select “Yes” from the drop-down menu if the pre-intervention measure was different from the post-intervention measure; select “No” otherwise.
  - If different forms of the same measure were used, it is not a different preintervention measure. If a different preintervention measure was used, consult with the review team leadership to determine whether the preintervention measure is an acceptable proxy for the post-intervention measure.
- **Did analysis adjust for pretest?** (column AV): Select “Yes” from the drop-down menu if the authors report an analysis that statistically controlled for the baseline difference; select “No” otherwise.

- **Okay to use in report?** (column AW): A formula will return “Yes” or “No” to indicate whether the measure should be included in the report.

Now, go back and enter a decision in cell B46 on the *Main Tab* (Are all of the following true?).

- Baseline equivalence is demonstrated for at least one analytic sample by calculating an RDD impact at the cutoff on the characteristics relevant to equating the groups as given in the protocol.
- This analysis used either exactly the same bandwidth and/or functional form as was used to estimate the impact on the outcome, or the same algorithm for selecting the bandwidth and/or functional form as was used to estimate the impact on the outcome.
- The statistical model used to estimate the impact on the outcome included statistical adjustment for characteristics relevant to equating the groups as given in the protocol, if needed.

### Analysis and results section of the data tab (columns AX-BE)

For each outcome that is eligible and *Meets WWC RDD Standards With or Without Reservations*:

- **Sample** (column AX): Indicate the sample for which intervention and comparison group means were computed. Possible options include “full sample” or a description of the subgroup or subsample for which means are provided (for example, “Grade 1” or “boys”).
- **Period** (column AY): Indicate any relevant information about the timing of the pre- or postintervention measure, such as when the preintervention variable was measured (for example, in the spring before random assignment or the school year) and whether it is a pretest only.

Reviewers should use either reported proportions (dichotomous outcomes) or reported covariate-adjusted means and SDs (continuous outcomes). Please consult review team leadership if you need assistance in determining how to estimate an effect size.

- **Posttest statistics**
  - **Dichotomous means:** Unadjusted postintervention means reported for both groups for variables that are true 0/1 variables (for example, dropout, graduated) and not for proportions (for example, percentage correct). Enter the means in column AZ (intervention) and column BC (comparison).
  - **Covariate-adjusted postintervention:** Covariate-adjusted postintervention means and full-sample standard deviations reported for both groups. Enter the means and SDs as follows: for the intervention group, columns BA and BB; for the comparison group, columns BD and BE.

### Findings section of the data tab (columns BF-BP)

For each outcome that is eligible and *Meets WWC RDD Standards With or Without Reservations*, complete columns BF-BP as appropriate.

- **N**(column BF): A locked cell with a formula to calculate the total sample size for some effect size calculations.
- **Study-reported effect size** (column BG): A cell for entering the study-calculated effect size.
- **S**(column BH): A locked cell with a formula to calculate the pooled SD.

- $g$ (column BI): A locked cell with a formula to calculate Hedges'  $g$  for continuous outcomes or Cox index for dichotomous outcomes.
- **Improvement index** (column BJ): A locked cell with a formula to calculate the improvement index based on the calculated Hedges'  $g$  or Cox index.
- **Study  $p$ -value** (column BK): Enter the  $p$ -value reported in the study.
- **Use study  $p$ -value?** (column BL): Select "Yes" from the drop-down menu if the WWC should use the  $p$ -value as reported in the study, such as when the author made the appropriate adjustments for baseline equivalence and clustering and conducted their analysis appropriately given their design; select "No" otherwise.
- **Did the study account for clustering?** (column BM).
  - If the unit of assignment and unit of analysis are the same, select "NA" from the drop-down menu.
  - If the unit of assignment and the unit of analysis are different and the study analysis is adjusted for this appropriately (such as by using hierarchical linear modeling), select "Yes" from the drop-down menu.
  - If the unit of assignment and unit of analysis are different and the study analysis did not adjust for this, select "No."
- **Multiple comparison rank** (column BN): Enter the rank of the outcome (smallest  $p$ -value = 1) if a multiple comparison correction is needed within the domain; leave the cell blank otherwise.
- **Critical  $p$ -value** (column BO): A locked cell with a formula to calculate the new critical  $p$ -value that will be used in determining statistical significance.

The following are brief instructions for determining statistical significance in the SRG. Refer to the Procedures Handbook for more details on the Benjamini-Hochberg (BH) correction procedures.

- To determine whether a finding is statistically significant after applying the BH correction, you should identify the  $p$ -value with the *largest* value that is statistically significant relative to its new critical  $p$ -value. Then all findings with a  $p$ -value less than that new critical  $p$ -value are statistically significant. This means you may identify as statistically significant a finding that was not significant relative to its own critical  $p$ -value.
- To assess statistical significance in the following example, start at the highest rank and compare the  $p$ -value of each outcome with the revised critical  $p$ -value based on its rank. In the example, the sixth-through eighth-ranked outcomes are not statistically significant. The fifth-ranked outcome is statistically significant ( $0.030 < 0.031$ ); therefore, this outcome and all others ranked higher (1-4) would be designated as statistically significant as well, including the fourth-ranked finding, in which the  $p$ -value is higher than its own revised critical  $p$ -value.

Author-reported or clustering-corrected $p$ -value ( $P_x$ )	Rank ( $x$ )	New Critical $p$ -value ( $P'x = 0.05x/X$ )	$P_x < P'x?$	Statistically significant after BH correction?
0.002	1	0.006	Yes	Significant
0.009	2	0.013	Yes	Significant
0.014	3	0.019	Yes	Significant
0.027	4	0.025	No	Significant
0.030	5	0.031	Yes	Significant
0.042	6	0.038	No	Not significant
0.052	7	0.044	No	Not significant
0.076	8	0.050	No	Not significant

- **Significant after multiple comparisons?** (column BP): Select “Yes” if the finding is statistically significant after adjusting for multiple comparisons; select “No” otherwise.

## SUMMARY TAB OF THE SRG

For Columns A-M, each row corresponds to a single-outcome domain.

- **Domain** (column A): Select a domain from the drop down list. This will be autopopulated from the domains entered in the *Data Tab*.
- **Number of outcomes** (column B): A locked cell with a formula to calculate the number of outcomes listed in the *Data Tab* that are in the domain entered in column A.
- **Effect size** (column C): A locked cell with a formula to calculate the average effect size for all outcomes in the domain entered in column A.
- **Max effect size** (column D): A locked cell with a formula to calculate the maximum effect size for all outcomes in the domain entered in column A.
- **Improvement index** (column E): A locked cell with a formula to calculate the improvement index from the average effect size for the domain entered in column A.
- **Min improvement index** (column F): A locked cell with a formula to calculate the minimum improvement index for any measure in the domain entered in column A.
- **Max improvement index** (column G): A locked cell with a formula to calculate the maximum improvement index for any measure in the domain entered in column A.
- **Max sample size** (column H): A locked cell with a formula to calculate the maximum sample size for any measure in the domain entered in column A.
- **Are any outcomes significant?** (column N): Select “Yes” or “No” from the drop-down menu.
- **Characterization of finding** (column M): Select the appropriate characterization of findings for the domain entered in column A. See the WWC Procedures Handbook for more details, but in brief, the following descriptions apply:

- **Single, SS+:** Estimated effect is positive and statistically significant after any necessary adjustments.
- **Single, indeterminate:** Estimated effect is neither statistically significant after any necessary adjustments nor substantively important.
- **Single, SS-:** Estimated effect is negative and statistically significant after any necessary adjustments.
- **Multiple, SS+ (A):** At least one main finding is positive and statistically significant, and none are negative and statistically significant based on univariate statistical tests, accounting for multiple comparisons and correcting for clustering when not properly aligned.
- **Multiple, SS+ (B):** The WWC-aggregated main finding is positive and statistically significant, correcting for clustering when not properly aligned.
- **Multiple, SS+ (C):** The study reports that the omnibus effect for all outcome measures together is positive and statistically significant based on a multivariate statistical test in a properly aligned analysis.
- **Multiple, indeterminate:** None of the main findings are statistically significant, accounting for multiple comparisons, and correcting for clustering when not properly aligned; the WWC-aggregated main finding is not statistically significant, correcting for clustering when not properly aligned.
- **Multiple, indeterminate:** At least one main finding is statistically significant and positive, and at least one main finding is statistically significant and negative, accounting for multiple comparisons and correcting for clustering when not properly aligned.
- **Multiple, SS- (A):** At least one finding is negative and statistically significant, and none are positive and statistically significant based on univariate statistical tests, accounting for multiple comparisons and correcting for clustering when not properly aligned.
- **Multiple, SS- (B):** The WWC-aggregated main finding for the multiple outcome measures is negative and statistically significant, correcting for clustering when not properly aligned.
- **Multiple, SS- (C):** The study reports that the omnibus effect for all outcome measures together is negative and statistically significant based on a multivariate statistical test in a properly aligned analysis.
- **Review information** (column N): These fields will autopopulate with information from the *Main Tab*.

## **AUTHOR QUERY & RESPONSE TAB OF THE SRG**

All reviewers should draft their questions for an author query (AQ). The MRG should capture the final set of questions as sent to the author. The MRG also should document whether (and when) a response was received and the response.

- **Date AQ sent** (row 1): Enter the date on which the AQ was sent in cell B1.
- **Response received?** (Yes/No) (row 2): Enter “Yes” or “No” in cell B2 to document whether any response was received.
- **Date of response** (row 3): Enter the date the response(s) was/were received by the team in cell B3.

Fill the questions and responses in rows 4-9, adding additional rows as required.

## REFERENCE

- Imbens, G., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of economic studies*, 79(3), 933-959. <https://doi.org/10.1093/restud/rdr043>
- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2), 698-714. <https://doi.org/10.1016/j.jeconom.2007.05.005>