



## Promoting Social and Behavioral Success for Learning in Elementary Schools

### Systematic Review Protocol

March 2022

Contract No. 91990019F0319

These practice recommendations were developed under Contract 91990019F0319, between 2M Research, 1521 N. Cooper St. Suite 600, Arlington, TX 76011, and the National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U. S. Department of Education. The views in this guide do not necessarily represent the views of the What Works Clearinghouse, the Institute of Education Sciences, or the U.S. Department of Education.

#### SUBMITTED TO:

National Center for Education Evaluation and Regional Assistance  
Institute of Education Sciences  
Department of Education  
550 12th Street SW  
Washington, DC 20202

Dr. Jonathan Jacobson  
[Jonathan.jacobson@ed.gov](mailto:Jonathan.jacobson@ed.gov)  
202-245-7485

#### POINT OF CONTACT:

Dr. Hiren Nisar  
[hnisar@2mresearch.com](mailto:hnisar@2mresearch.com)  
703-214-1211

2M Research  
1521 N. Cooper St. Suite 600  
Arlington, TX 76011  
[www.2mresearch.com](http://www.2mresearch.com)

# Promoting Social and Behavioral Success for Learning in Elementary Schools

## SYSTEMATIC REVIEW PROTOCOL

**March 2022**

Hiren Nisar  
Dallas Elgin  
**2M Research**  
1521 N. Cooper St. Suite 600  
Arlington, TX 76011

Expert Panel

Catherine Bradshaw (Chair)  
*University of Virginia*

Virginia Dolan  
*Retired, Formerly of Anne Arundel County  
Public Schools*

Andy Frey  
*University of Louisville*

Rob Horner  
*University of Oregon*

Julie Owens  
*Ohio University*

Kelly Perales  
*Midwest Positive Behavioral Interventions  
and Supports Network*

Kevin Sutherland  
*Virginia Commonwealth University*

# SYSTEMATIC REVIEW PROTOCOL: PROMOTING SOCIAL AND BEHAVIORAL SUCCESS FOR LEARNING IN ELEMENTARY SCHOOLS

## VERSION 3.0 (MAY 2020, REVISED JANUARY 2022)

---

This protocol guides the systematic review of research that informs recommendations for the guide, *Promoting Social and Behavioral Success for Learning in Elementary Schools*. The protocol is used in conjunction with the [What Works Clearinghouse Procedures and Standards Handbooks \(Version 4.0\)](#).

### Purpose Statement

The What Works Clearinghouse (WWC) produces practice guides that offer recommendations backed by credible research. *The Reducing Behavior Problems in the Elementary School Classroom* practice guide, published in 2008, offers five recommendations pertaining to classroom behavior management. Although these recommendations remain relevant today, since their publication there have been more than 10 years of policy development, practical experience, and research rooted in topics like tiered instruction, social-emotional learning, positive and prosocial behavior, prevention, and classroom management. Therefore, this practice guide will focus on recent research to generate updated recommendations that will assist practitioners with their efforts to promote positive behavior for all students in classrooms and schools.

Positive behavior entails observable actions by students that promote academic and/or social success for students, peers, and faculty in schools. Positive behavior may include actions such as following classroom rules and expectations, self-regulation, helping others, playing or working cooperatively, and attending to instruction.

Whole-school or whole-class programs designed to promote positive behavior can enhance the likelihood that effective teaching and learning occur for all students (Benner, Nelson, Sanders, & Ralston, 2012; Bradshaw, Waasdorp, & Leaf, 2012; Gietz & McIntosh, 2014). This practice guide is intended to help elementary education educators, as well as school and district administrators and parents, implement and support effective prevention-focused strategies that promote positive student behavior and related outcomes. In considering the growing body of research on prevention-focused frameworks to improve students' academic and behavior outcomes, this practice guide focuses on whole-school or whole-class practices to support positive behavior by elementary school students, including students identified as having disabilities or who are at risk of identification for special education services.

## Research Questions

The following research questions motivate this review:

- Which whole-class or whole-school (Tier 1) practices are effective at:
  - supporting positive behavior among students in elementary school?
  - supporting change in educators' instructional and discipline practice?
  - improving the school or classroom environment, including measures of school climate and equity?
- Are some practices especially effective for students identified with disabilities or who are at risk of identification for special education services while learning in a Tier 1 context?
- What supports are needed to help educators implement these practices with fidelity to promote positive behavior?

## Key Definitions

**Positive behavior.** Positive behavior includes observable actions such as self-management—including self-monitoring, self-evaluation, self-delivered consequences, and self-recruited support; social interactions; or other behaviors that yield social and/or instructional benefits in a school setting.

**Prosocial behavior.** Prosocial behavior is closely related to positive behavior, but for the purposes of this review, prosocial behavior refers to observable actions characterized by *interacting with others*, including peers and school staff, and behaving in ways to benefit other people. Common examples of prosocial behavior include sharing, volunteering, and protecting someone else from bullying. In the context of the classroom and school environment, students are presented with both formal instruction such as collaborative learning activities, and informal opportunities such as interactions with peer groups, relationships with teachers, and play to develop skills that support prosocial and positive behavior. Aspects of prosocial skill development that might be targeted for this review include those related to underlying social competencies such as emotional regulation, perspective-taking, and cognitive and social problem-solving.

**School climate.** School climate refers to the environment of the school, or classrooms within a school. The National Center on Safe Supportive Learning Environments (n.d.) defines school environment as a school's "facilities, classrooms, school-based health supports, and disciplinary policies and practices." Aspects of the school environment include the physical environment—such as reducing overcrowding and messaging positive behavior expectations on posters; the relational environment—such as improving peer-to-peer communication and reducing bullying behavior; and the instructional environment—such as improving communication practices and approaches between students and staff.

**School equity.** School equity refers to the policies and practices of the school, or classrooms within a school, that promote an equal chance of success for all students. School equity can be understood in terms of (1) fairness, whereby personal or social circumstances do not prevent students from achieving their academic potential and (2) inclusion, whereby a basic minimum standard is shared by all students regardless of personal characteristics (OECD, 2012). School and classroom equity deal with matters such as the disproportionality of discipline practices across different racial, ethnic, and special education subgroups. Typical assessments of equity entail assessing the rate of disciplinary practices such as suspensions across two groups of students within a school, or student or staff surveys of their perceptions of equity.

## **Eligibility Criteria**

### **Eligible Populations**

Studies that examine the effects of interventions on positive behavior within whole classrooms and schools are eligible for this review. Students must be in elementary school. This generally means students in grades K–6 or ages 5–11 when the intervention is implemented. When schools include higher grades, such as K–8, related outcomes may be included if they cannot be disaggregated to include only grades K–6. Schools with a grade 6–8 or higher grade structure represent a population that is ineligible for this review.

Students with disabilities may be included in the sample so long as the intervention is also intended to be used with, and meets the needs of, students without disabilities.

Studies that examine the effects of interventions on teachers and, more broadly, on educators, can also be included in this review. Hence, elementary school teachers or other school staff members represent eligible populations.

### **Eligibility of Findings from Multiple Analyses in a Study**

This review follows the guidance in the *WWC Procedures Handbook* (in Chapter IV: Reporting on Findings) regarding reporting on findings from subgroups, multiple analyses that use composite or subscale scores, or different time periods. The WWC reports findings from all eligible analyses that meet standards, split into main and supplemental findings. Main findings will be reported in the body of the practice guide if they are used for practice guide recommendations. Other eligible findings that meet standards can be included in supplemental appendices to the practice guide or can be reported on the WWC website but will not be prioritized for review. For each outcome measure, and among those findings that meet WWC design standards, WWC uses the following criteria to designate those findings as the main ones to report: (1) includes the full sample; (2) uses the most aggregate measure of the outcome measure, rather than individual subscales or narrower constructs within the same domain; and (3) is measured at a time specified by the protocol.

Under this review, findings for the potential subgroups of interest listed in Table 1 are also eligible to be reported in supplemental appendices in the practice guide or on the WWC website if review team leadership decides to review these findings and the WWC publishes them online. Findings for other subgroups are not eligible for review unless designated as the main finding based on the criteria above.

There may be studies with multiple outcomes classified within the same domain. On these occasions the review team leadership will identify the key outcomes to be considered in subsequent analyses. This determination will be based on sample size, breadth of the measure, correspondence with similar outcomes in other studies, and alignment with the presentation of outcomes in the study. Outcomes in the domain not identified as a main outcome (such as subscales) could, at the discretion of review team leadership, be reviewed and reported as supplemental findings in appendices to the practice guide.

For this review, measures obtained at the end of an intervention, as well as at any point, thereafter, are admissible. When reported, this review will classify findings for outcomes administered immediately after the intervention—for example, outcomes administered after the third year of a 3-year intervention is completed—as main findings, because these findings are more prevalent in the studies eligible for review for this practice guide. Measures occurring several months after the intervention may also provide strong evidence for an intervention’s effectiveness. Additionally, intermediate outcome measures that reflect partial exposure to an intervention can provide useful information about the intervention’s effectiveness. Therefore, at the discretion of review team leadership, follow-up and intermediate findings, when available and appropriate, may be reported in appendices to the practice guide or on the WWC website.

**Table 1. Subgroups of Interest to the Supporting Positive Behavior Practice Guide**

Characteristics of Students	Characteristics of Setting or Context	Characteristics of Educators
<ul style="list-style-type: none"> <li>• Age or grade level</li> <li>• Gender</li> <li>• Economically disadvantaged—for example, free or reduced-price lunch eligibility, low-income status, or parental education</li> <li>• Race or ethnicity</li> <li>• English learner status</li> <li>• Disability status or disability category</li> </ul>	<ul style="list-style-type: none"> <li>• Location of the schools involved—for example, urban, suburban, rural</li> <li>• School governance—for example, regular public, charter, private, religious</li> <li>• Economically disadvantaged school—for example, Title I status</li> </ul>	<ul style="list-style-type: none"> <li>• Position—for example, paraprofessional, teacher, administrator</li> <li>• Years of experience</li> <li>• General or Special Education Teacher</li> </ul>

While the above rules will guide how we identify main and supplemental findings, review team leadership has discretion to identify main and supplemental findings after considering additional

factors about the findings under review, such as the prevalence of findings across implementation levels and the design of the intervention.

## Eligible Interventions

The term *intervention* as used throughout this protocol refers to (1) a framework, (2) a program or combination of program elements, or (3) practices or strategies designed to produce behavior changes. In order to be eligible for review for this practice guide, the intervention must in some way seek to improve students' positive behaviors. It is equally permissible to review interventions that seek to stop problematic behavior, as defined by study authors. To be eligible, interventions must be designed to be used at whole-classroom or whole-school levels.

Interventions used at different units of analysis—for example, groups of students within a classroom, or individual students—are eligible for review only if they could be delivered to a whole elementary school or classroom, or in support of a whole elementary school or classroom.

To be eligible, the intervention must be delivered in a school setting. Public, private, and charter schools are all eligible settings. Studies in alternative school settings or education facilities with unusual levels of support or instructional approaches are not eligible for review. A residential treatment facility is the primary example of an ineligible setting. Eligible interventions or programs may include, for example, community or parental activities, if there is also a school-based component.

Practice guides recommendations must be feasible to implement in a typical classroom. To that end, only interventions that are reasonably replicable are eligible to review. The review team will consider the following factors in determining the potential replicability of interventions:

- Intervention description: skills being targeted; approach to enhancing the skill(s)—for example, strategies, activities, and materials; unit of delivery of the intervention, such as a whole group; medium/media of delivery—for example, teacher-led instruction or software; and targeted population
- Intervention duration and intensity
- Description of individuals delivering or administering the intervention

In this review, the following types of interventions may be included:

- ***Practices and strategies.*** The review includes schoolwide or classroom-level practices intended to promote positive behavior. For example, “Positive Behavioral Interventions and Supports” is a set of practices/strategies that would be eligible for review.
- ***Programs or services.*** The review includes programs or services that can be purchased or acquired for school or classroom use to support positive behavior, including training for school staff, curricula, or other materials. For example, the “School Climate Improvement Resource Package” would be eligible for review.

- **Policies.** The review includes schoolwide or classroom-level policies, such as enacting an antibullying policy, intended to promote a positive learning environment and/or accentuate the importance of students’ positive development and positive behavior.

An eligible positive behavior intervention may include providing professional development to support staff who are delivering the intervention. This review will describe the professional development provided to staff delivering the intervention based on the information reported in the studies.

Review team leadership will give priority to reviewing eligible research on *nonbranded* interventions, with the goal of informing the development of practice recommendations that are not specific to a particular brand of intervention. *Branded* interventions are commercial or published programs and products that may possess any of the following characteristics:

- An external developer who provides technical assistance (for example, instructions/guidance on the implementation of the intervention) or sells or distributes the intervention
- Trademark or copyright

Although this review will describe academic outcomes, interventions with an exclusive focus on academic improvement will be ineligible for review. For example, a professional development intervention solely intended to improve teacher academic instruction will not be eligible for review for this practice guide. However, if the professional development is also specifically intended to change aspects of student behavior, then a study of the intervention would potentially be eligible for review.

## Eligible Research

The *WWC Procedures Handbook* discusses the types of research reviewed by WWC in “Section II: Developing the Review Protocol” and “Section III: Identifying Relevant Literature.” Additionally, in this review, the following additional parameters define the scope of research studies to be included:

- **Timeframe.** For this practice guide, which is an update to a practice guide published in 2008, the study must have been released in 2007 or later. Studies must be publicly available—that is, accessible on the web or available through a publication, such as a journal—at the time of the original or updated literature search.
- **Sample.** The study sample must meet the requirements specified above in the *Eligible Populations* section at the time they receive the intervention. For example, while the students in the sample must be in elementary school at the time they receive intervention, their outcomes can be measured after they advance to later grades.
- **Language.** The study must be available in English to be included in the review.



- **Location.** The study must include students in the United States, in its territories or tribal entities, at U.S. military bases overseas, or in other Organization for Economic Cooperation and Development member countries in which English is the primary or most commonly used language. This includes Australia, Canada, Ireland, New Zealand, and the United Kingdom.

## Eligible Outcomes

There are 10 outcome domains specified for this review, organized within 3 categories: (1) Student Outcomes, (2) Educator Outcomes, and (3) Environmental Outcomes. Table 2 lists the positive behavior outcome domains that fall under each of the three categories. Unlike the domains themselves, the categories are for descriptive purposes only and do not influence how the outcomes are reviewed or reported in the practice guide or on the WWC website.

**Table 2. Outcome Domains Focused on Students, Educators, and the Environment**

Student Outcomes	Educator Outcomes	School Environment Outcomes
<ul style="list-style-type: none"> <li>• Student Engagement in School</li> <li>• Compliant Student Behavior</li> <li>• Student Social-Emotional Functioning</li> <li>• Primary School Academic Achievement</li> <li>• General Literacy Achievement</li> <li>• General Mathematics Achievement</li> </ul>	<ul style="list-style-type: none"> <li>• Educator Discipline Practice</li> <li>• Educator Instructional Practice</li> </ul>	<ul style="list-style-type: none"> <li>• Classroom and School Climate</li> <li>• Classroom and School Equity</li> </ul>

These outcomes may include those measured using surveys; administrative records; systematic observations of teaching practices; and the perspectives of students, teachers, researchers, observers, school staff, and parents.

Measures based on self-reports are not eligible when respondents who provide the data are aware of their study condition. As an example, the “Internalizing Problems Subscale” of the “Behavior Assessment Scale for Children” includes teacher and parent reports and student self-reports. The teacher reports would only be eligible if completed by someone besides the teacher participating in the intervention. However, parent reports and student self-reports would potentially be eligible.

In the context of this review, the concern about self-reported outcomes is primarily restricted to teacher reports. Furthermore, measures based on teacher self-reports can be eligible for review, even when respondents who provide the data are aware of their study condition, in studies where the tested intervention is part of a district or statewide initiative. In these cases, it can be possible that teachers are not expected to require intervention “buy-in” to the point that there is reason to

be concerned that their reporting about student behavior is likely to be biased. Review team leadership has discretion to render a judgment for studies that fit these criteria.

Following is a description of each outcome domain, within each category.

### ***Student Outcomes***

***Student Engagement in School.*** Student engagement in school is demonstrated through positive behaviors that can be observed during academic tasks or events that influence the degree to which students can be engaged in an academic task. Engagement is also demonstrated when students put forth effort to be successful at school. Examples include

- staying on task during a class assignment,
- coming to class prepared,
- completing assignments (including homework assignments),
- following school or classroom rule, and
- participating in school activities.

This domain captures concerns about school engagement, such as the above examples but with negative valences. Examples include

- inattention and
- lack of engagement.

Constructs that fall within this domain will generally be those that occur in the school environment but need not actually occur within school, such as homework completion rates. Note that academic achievement, such as performance or mastery of specific academic skills, is captured in student academic outcome domains described below. Additionally, disruptive behavior that occurs in a classroom should fall under the “Compliant Student Behavior” domain.

***Compliant Student Behavior.*** This domain primarily includes observed behaviors that have individual consequences for the student. Compliant student behavior also includes behavior that does not have a direct connection to academic effort. Examples include

- disruptive behavior including talking out loud at inappropriate times;
- delinquent behaviors such as lying, cheating, or stealing;
- physical aggression such as fighting;
- absenteeism and tardiness;
- impulsivity;
- acts of bullying;
- office discipline referrals (ODRs); and
- school suspension or expulsion.

A key differentiating factor between outcomes in the “Compliant Student Behavior” domain and the “Student Social-Emotional Functioning” domain is that the latter domain entails interactions with others. This is generally not true of outcomes in the “Compliant Student Behavior” domain; however, interactions that take place at inappropriate times would fall under this domain.

Eligible behaviors in this domain can occur in or away from school. For example, parental reports of disruptive behavior at home are eligible for review.

***Student Social-Emotional Functioning.*** This domain includes student behaviors and self-ratings that primarily reflect a student’s emotional state and/or ability to effectively manage relationships with others. Social-emotional functioning measures may be based on a diagnosis or classification, student self-report, teacher observation—such as a report on the degree to which students seem emotionally withdrawn—or results from an assessment scale. Examples include

- self-awareness of thoughts, feelings, and behaviors;
- ratings of adaptive functioning;
- measures of self-control/self-regulation/self-management;
- overall adjustment/wellbeing;
- emotional regulation;
- thought disorders;
- depression; and
- anxiety.

This domain also captures student social interactions. Examples include

- peer rejection;
- isolation from others; and
- demonstrating specific social skills, such as social awareness of context and others and interpersonal relationship skills.

***Primary School Academic Achievement.*** Primary school academic achievement includes academic measures based on student test scores across multiple subjects, including at least two of the following subjects: literacy, mathematics, science, and social studies. It pertains only to students in elementary school or to learners ages 5–11. This domain should only be selected for test score measures that *do not* correspond with a more specific domain.

***General Literacy Achievement.*** General literacy achievement includes two or more of the following literacy domains: alphabetics, comprehension, reading fluency, writing conventions, writing productivity, and writing quality.

***General Mathematics Achievement.*** General mathematics achievement includes content in two or more of the following mathematical domains, as well as tests of mathematical understanding, procedures, and problem solving designed to measure more than one content area: algebra; data

analysis, statistics, and probability; geometry and measurement; numbers and operations; calculus; and trigonometry/precalculus.

### ***Educator Outcomes***

Measures based on teachers' observations or self-reports of their own actions are not eligible when teachers who provide these types of data participated in the intervention. This is because teachers might be influenced by knowing their study condition.

***Educator Discipline Practice.*** Educator discipline practices are attempts to influence problematic student behavior by responding to students' actions with consequences or rewards. In contrast, practices that focus more on *prevention* of problematic behavior (such as classroom management practices not solely focused on consequences and rewards) would instead potentially be eligible under the "Educator Instructional Practice" domain. Eligible outcomes in the "Educator Discipline Practice" domain include, but are not limited to, the following:

- The number of ODRs made by a teacher
- Relevant subscales about discipline practice from measures based on teacher self-reports, classroom observations, or student surveys
- Use of behavior-specific praise

***Educator Instructional Practice.*** Educator instructional practices include measures that reflect the quality of instruction provided by teachers and their application of content knowledge or pedagogical content knowledge as demonstrated by their actions in the classroom. These measures can be based on rubrics assessed by school principals, supervisors, or trained evaluators, or can be based on surveys administered to students. Eligible assessments of the quality of educator instructional practice include, but are not limited to, the following:

- Charlotte Danielson's Framework for Teaching (FFT)
- Classroom Assessment Scoring System (CLASS™)
- Protocol for Language Arts Teaching Observations (PLATO)
- Mathematical Quality of Instruction (MQI, predicting mathematics achievement)
- Tripod
- UTeach Teacher Observation Protocol (UTOP)

This review will focus on educator instructional practice measures to the extent that the intervention is thought to influence student outcomes, including student achievement and other social-emotional and behavior outcomes. For this reason, as described in the section on outcome measure requirements below, eligible measures in the "Educator Instructional Practice" domain must satisfy a validity requirement with a statistical relationship between the outcome measure and student achievement, social-emotional learning, or behavior. All measures of instruction named above meet this validity requirement for student achievement based on evidence reported in the Measures of Effective Teaching study (Kane & Staiger 2012).

## ***School Environment Outcomes***

***Classroom and School Climate.*** The “Classroom and School Climate” domain includes measures that describe the environment of the classroom or school and can be filled out by, for example, students and teachers. Eligible measures in this domain include observations or assessments of the classroom/school environment, the quality of interpersonal relationships within the classroom, and other factors that describe the character of a school. Measures that describe the characteristics of an individual staff member or the characteristics of students are not eligible in this domain. While measures of equity could be considered part of a classroom or school climate, for this review measures of equity will instead potentially be eligible under the “Classroom and School Equity” domain.

***Classroom and School Equity.*** The “Classroom and School Equity” domain includes measures of the degree of equity or assessments of equity within classrooms or schools. Eligible measures in this domain include disproportionality assessments (for example, whether the rate of ODRs or suspensions differs between two groups of students within a school) and student or staff surveys of their perceptions of equity. Keep in mind that measures based on self-report in instances where respondents are aware of the treatment condition are not eligible.

## **Evidence Standards**

Eligible studies are assessed against WWC evidence standards, as described in the *WWC Procedures Handbook*, “Section IV: Screening Studies” and “Section V: Reviewing Studies,” as well as the *WWC Standards Handbook*.

## **Sample Attrition**

The *WWC Standards Handbook* discusses the sample attrition standards used by WWC in the following sections:

- Step 2 of the WWC review process for individual-level group design studies in Section II.A—Sample Attrition: Is the combination of overall and differential attrition high?
- Step 1 of the WWC review process for cluster-level group design studies in Section II.B—Is the study a cluster RCT with low cluster-level attrition?
- Step 3 of the WWC review process for cluster-level group design studies in Section II.B—Is there a risk of bias due to non-response of individuals?
- Section 3 of the WWC Standards for reviewing complier average causal effect (CACE) estimates in Section II.D—Calculating attrition when rating CACE estimates.
- Standard 2 of the WWC Standards for reviewing regression discontinuity designs in Section III.C.

This review uses the *optimistic* boundary for attrition. This boundary assumes that attrition in studies of positive behavior interventions within a K–6 setting is not especially likely to be due

to factors strongly related to intervention status. Younger students are, compared to older students, less likely to have autonomy with respect to school attendance and decisions around school transfer. Standard concerns around family mobility and teacher turnover associated with intervention status should be accounted for by the optimistic boundary.

In the *WWC Standards Handbook*, Figure II.2 illustrates the attrition boundary, and Table II.1 reports attrition levels that define high and low attrition. The study review guide calculates attrition and characterizes whether it is high or low based on these rules.

## Joiners in Cluster Randomized Controlled Trials

WWC defines a *joiner* as any student who enters a cluster (for example, a school or classroom) after the results of random assignment are known to any individual who could influence a student's placement into a cluster (for example, parents, students, teachers, principals, or other school staff). The presence of joiners in an analytic sample has the potential to introduce bias into estimates of an intervention's effectiveness. Note: Given that educator outcomes are also of interest for this practice guide, teachers can also be considered as potential joiners in some studies.

In some cases, joiners who enter clusters relatively early in the study period have less potential to introduce bias than those who enter later. Therefore, WWC sometimes differentiates between *early joiners* and *late joiners*. For this review protocol, students will be considered *early joiners* if they enter a cluster within a 6-week time frame after the results of random assignment are known or, in cases for which random assignment occurred during the summer, 6 weeks after the start of the school year. *Late joiners* are those students who enter clusters after the end of the early joiner period.

This review protocol specifies that for cluster randomized controlled trials (RCTs) for which the unit of assignment is smaller than a school, *all joiners in the analytic sample pose a risk of bias*. This fact holds true because classroom rosters are often determined by school administrators who might assign students to classrooms based on knowledge of the intervention. Additionally, students or parents may influence assignment to within-school clusters (for example, classrooms) because they may have a specific preference for or against the intervention. Therefore, a study that includes at least one such joiner in the analytic sample and uses within-school clusters *does not limit the risk of bias from joiners* (whether a joiner is early or not).

For cluster RCTs of schoolwide interventions for which the unit of assignment is a school or larger cluster (such as a district), whether early joiners pose a risk of bias depends on details about the intervention and, in particular, whether the intervention is expected to have influenced school enrollment or placement decisions. The two most common examples for this practice guide are described below:

- If the intervention is unlikely to directly affect enrollment or placement decisions, then ***only late joiners pose a risk of bias***. Again, joiners who enter clusters after 6 weeks will be considered late joiners. Even if it seems likely that an intervention would not directly affect placement decisions, late joiners are still considered to be a potential source of bias. A study of an intervention that involved school-level cluster assignment that included at least one late joiner in the analytic sample *does not limit the risk of bias from joiners*.
- If the intervention may affect enrollment or placement decisions (such as a highly publicized program to reduce school bullying in a high-risk area), then ***all joiners pose a risk of bias***. A study of such an intervention that includes one or more joiners in the analytic sample, whether early or not, *does not limit the risk of bias from joiners*.

For studies reviewed for this practice guide, the default assumption is that the interventions being examined are unlikely to affect enrollment or placement decisions; however, review team leadership has discretion to revise this guidance. Typical scenarios WWC encounters in cluster RCTs are described above, but not all scenarios can be anticipated. When an intervention and unit of assignment in a cluster RCT do not fall into a category described above, review team leadership has discretion to decide whether the joiners pose a risk of bias.

## Baseline Equivalence

If the study design is an RCT or regression discontinuity design (RDD) with high levels of attrition or a quasi-experimental design, the study must satisfy the baseline equivalence requirement for the analytic intervention and comparison groups. The *WWC Standards Handbook* discusses how authors must satisfy the baseline equivalence requirement in the following sections:

- Step 3 of the WWC review process for individual-level group design studies in Section II.A—Baseline Equivalence: Is equivalence established at baseline for the groups in the analytic sample?
- Steps 4 and 7, respectively, of the WWC review process for cluster-level group design studies in Section II.B—Does the study establish equivalence of individuals at baseline for groups in the analytic sample? and Does the study establish equivalence of clusters at baseline for groups in the analytic sample?
- Section 5 of the WWC Standards for reviewing CACE estimates in Section II.D—Procedures for Rating CACE Estimates when Attrition is High.
- Standard 3 of the WWC Standards for reviewing regression discontinuity designs in Section III.C.

### ***Baseline Equivalence of Individuals***

For studies that must satisfy baseline equivalence of individuals, including cluster-level assignment studies being reviewed for evidence of effects on individuals, the baseline

equivalence requirement must be satisfied for the analytic intervention and comparison groups. Baseline differences for one analytic sample do not affect the assessment of baseline equivalence for a finding that uses a separate analytic sample. For this review, baseline equivalence must be satisfied on one of the following preintervention (or baseline) characteristics:

1. A preintervention measure of the outcome used in the analysis, when available. This approach should be used first by reviewers.
2. A preintervention measure of a different variable from within the same domain as the outcome used in the analysis. For example, a preintervention measure of ODRs could be used to show baseline equivalence of an analysis of school suspensions, since both ODRs and school suspensions are found in the “Observed Individual Behavior” domain. If the first approach is not possible, reviewers should use this approach.
3. Two preintervention measures from within two other distinct student outcome domains. This third approach is available only for outcome measures that fall in one of the seven domains within the “Student Outcomes” domain (see the list in Table 2). For example, one preintervention measure in the “General Literacy Achievement” domain and a second in the “Student Engagement in School” domain can be used to satisfy the baseline equivalence requirement for outcomes in the “Observed Individual Behavior” domain. This approach should only be used if the baseline equivalence requirement cannot be satisfied using either of the previous two approaches.

While the study must use one of these three approaches to determine whether the baseline equivalence requirement can be satisfied, WWC will also consider the size of baseline differences for an analytic sample within a given domain. First, if the analytic sample for a study finding that must satisfy the baseline equivalence requirement has a baseline effect size greater than 0.25 standard deviations for any preintervention measure in a given domain, all findings in that domain for that analytic sample will be rated *Does Not Meet WWC Group Design Standards*. In addition, for student outcomes, WWC will assess the baseline difference between intervention and comparison groups for each eligible preintervention measure in a student outcome domain. If the baseline differences are greater than 0.25 standard deviations for any two preintervention measures from within two distinct student outcome domains, all findings for student outcomes using that analytic sample will be rated *Does Not Meet WWC Group Design Standards*.

If the analytic sample for a study finding that must satisfy the baseline equivalence requirement has a baseline effect size between 0.05 and 0.25 standard deviations (so that a statistical adjustment is required) for *any* preintervention measure within the domain, all outcomes within that domain must adjust for that baseline difference. For example, if A, B, and C are available as pre- and post-intervention measures from the same outcome domain, and the preintervention difference for B requires statistical adjustment, then the impact analyses for A, B, and C must all be adjusted for B to be eligible to meet WWC standards with reservations. Otherwise, the findings in this domain are rated *Does Not Meet WWC Group Design Standards*. Note that a statistical adjustment is required only within the domain, and there is no need to adjust for differences across domains, provided that baseline equivalence is established using one of the



first two approaches listed above. When the third approach is used to establish baseline equivalence for a student outcome (using preintervention measures from two other student outcome domains), the analysis must include statistical adjustments for all preintervention measures within those two student outcome domains with baseline differences in the adjustment range.

In addition to the preintervention measures required for satisfying the baseline equivalence requirement, a large baseline difference on some other characteristic could be evidence that the intervention and comparison groups are not sufficiently comparable for the purposes of the review. For study reviews for this practice guide, these characteristics include measures of students' age and grade level. When differences in student age or grade level are larger than 0.25 standard deviations, the study finding will be rated *Does Not Meet WWC Group Design Standards*. If the study does not report these characteristics but describes a study sample that gives the reviewer reason to question the magnitude of the differences on these characteristics, review team leadership has discretion to conduct an author query to obtain information on the similarity of the groups on age and grade level.

### ***Baseline Equivalence of Clusters***

***Assessing equivalence of clusters.*** In general, considerations for satisfying baseline equivalence of individuals also apply to satisfying baseline equivalence of clusters. In particular, baseline equivalence of clusters in the intervention and comparison groups must be satisfied by one of the same baseline measures described above for assessing baseline equivalence of individuals, and the same statistical adjustment requirements apply.

***Acceptable baseline time periods for establishing equivalence.*** For the positive behavior review, any of the following three sources of baseline data can be used to satisfy the baseline equivalence requirement for the analytic sample of clusters (provided the data are representative of the individuals who were within the clusters at the time the baseline data were collected):

1. The analytic sample of individuals from any preintervention time period.
2. Individuals from the same cohort and within the same clusters as the individuals in the analytic sample. The baseline data may be obtained at the time that clusters were assigned to conditions or during the school year prior to when clusters were assigned to conditions.
3. Individuals from the previous (adjacent) cohort, in the same grade, and within the same clusters as individuals in the analytic sample.

If authors provide baseline information at multiple time periods, a reviewer should assess baseline equivalence using the information collected at the period closest to the start of the intervention. If authors provide baseline information for multiple samples—for example, the analytic sample and an adjacent cohort—a reviewer should assess baseline equivalence using the sample listed first above; that is, (a) should be used if available, then (b), and then (c). If authors

provide baseline information for multiple samples across multiple time periods, the reviewer should consult with the review team to determine which information to prioritize.

When a study examines the effectiveness of an intervention in multiple time periods, the sample used to satisfy baseline equivalence of clusters in the base period (for example, the school year after random assignment) also satisfies baseline equivalence of clusters in the later time periods (for example, 2 years after random assignment), so long as the outcome data are representative of the individuals in the clusters.

## **Outcome Measure Requirements**

In this review the requirements for outcome measures, including reliability of outcomes, follow those specified in the *WWC Standards Handbook* (in Section “IV.A: Outcome Requirements and Reporting”). As described in the *WWC Standards Handbook*, outcomes based on administrative records (such as number of ODRs or attendance) will generally be considered to meet reliability requirements.

## **Statistical Adjustments**

The *WWC Procedures Handbook* discusses the types of adjustments made by WWC in “Section VI: Reporting on Findings.” For *mismatched* analysis (that is, when a study assigns units at the cluster level but conducts analysis at the individual level), reviews for this practice guide use the intra-class correlation coefficient of 0.20 for the achievement domain and 0.10 for all other domains, unless a study-reported intra-class correlation coefficient is available.

## **Eligible Study Designs**

Studies that use group designs (RCTs or quasi-experimental designs) are eligible for review using the appropriate WWC design standards. Studies utilizing RDDs or single-case designs (SCDs) may warrant additional consideration and therefore should be referred to the review team leadership to determine their prioritization for review.

## **Procedures for Conducting the Literature Search**

The *WWC Procedures Handbook* discusses the procedures for conducting a literature search in “Section III: Identifying Relevant Literature” and “Appendix B: Policies for Searching Studies for Review.” This first step of conducting the literature review will use a quick literature search process to identify search terms that may be relevant for promoting social and behavioral success for learning in elementary schools practice guide using broad keywords. In this step, content experts will identify and recommend broad keyword search terms related to promoting positive and prosocial behavior. This process will be further narrowed, depending on the number of studies found, to remove some of the generic terms. The initial review of literature may also identify additional search terms that may be included in our search strategy.

The second step of the process is to conduct intervention-specific literature searches, using the intervention name, to identify all publications on interventions related to this practice guide. The review of studies from this intervention-specific search may refine the potential scope of the final search by including additional terms in the final list of search terms. For example, search terms applicable for the positive behavior area include phrases such as school climate, antibullying, and nonacademic outcomes, among others. We will finalize the search terms related to broad keywords, intervention-specific keywords, population of interest-specific keywords, outcomes-specific keywords, and design-specific keywords. These well-specified search terms will identify the studies from a set of electronic databases relevant for this practice guide (see Table B.2 in Appendix B of the *WWC Procedures Handbook*).

In a third step, each citation gathered through this literature search will undergo a screening process to determine whether the study meets the eligibility criteria established in the review protocol. This screening process is described in Chapter IV of the *WWC Procedures Handbook*. Finally, the studies are prioritized for review based on the study potentially meeting WWC standards (with or without reservations). This prioritization process is described in Appendix A of the *WWC Procedures Handbook*.

## **Additional Sources**

Literature reviews for this practice guide involve searching the electronic databases listed in Appendix B of the *WWC Procedures Handbook*, as well as searching the following websites:

- Academy for Social-Emotional Learning in Schools
- American Academy of Pediatrics
- American Educational Research Association
- American Psychological Association
- ASCD: Professional Learning & Community for Educators
- Association for Positive Behavior Support
- Blueprints for Healthy Youth Development
- Centers for Disease Control and Prevention
- Children’s Hospital of Philadelphia
- Channing Bete Company
- Collaborative for Academic, Social, and Emotional Learning (CASEL)
- Committee for Children
- Council for Exceptional Children
- Edutopia
- Hazelden
- National Association of Elementary School Principals
- National Association of School Psychologists
- National Association of Secondary School Principals
- National Association of Special Education Teachers

- National Center on Safe and Supportive Learning Environments
- National Clearinghouse on Supportive School Discipline
- National Education Association
- National Institutes of Justice
- National School Climate Center
- Positive Behavior Intervention Supports
- School-Counselor.org
- Stopbullying
- The School Superintendent Association

## References

- Benner, G. J., Nelson, R., Sanders, E. A., & Ralston, N. C. (2012). Behavior interventions for students with externalizing behavior problems: Primary-level standard protocol. *Exceptional Children, 78*(2), 181–198. doi: 10.1177/001440291207800203
- Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2012). Effects of school-wide positive behavioral interventions and supports on children with behavior problems. *Pediatrics, 130*(5), e1136–e1145. doi: 10.1542/peds.2012-0243
- Gietz, C. & McIntosh, K. (2014). Relations between students’ perceptions of their school environment and academic achievement. *Canadian Journal of School Psychology, 29*, 161–176. doi: 10.1177/0829573514540415
- Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle, WA: Bill & Melinda Gates Foundation. Retrieved from <http://k12education.gatesfoundation.org/resource/gathering-feedback-on-teaching-combining-high-quality-observations-with-student-surveys-and-achievement-gains-2/>.
- National Center on Safe Supportive Learning Environments. (n.d.). *Safe and healthy students: Environment*. Retrieved from <https://safesupportivelearning.ed.gov/topic-research/environment>.
- OECD. (2012). *Equity and quality in education: Supporting disadvantaged students and schools*. OECD Publishing. <https://doi.org/10.1787/9789264130852-en>