

REVIEW PROTOCOL FOR SUPPORTIVE LEARNING ENVIRONMENT INTERVENTIONS VERSION 4.0 (MAY 2020)

This protocol guides the review of research that informs the What Works Clearinghouse (WWC) intervention reports in the Supportive Learning Environment topic area. The protocol is used in conjunction with the *WWC Procedures and Standards Handbooks (Version 4.0)*.

PURPOSE STATEMENT

A supportive learning environment is one that is safe, includes effective and fair use of appropriate discipline practices, and promotes positive student outcomes and teacher practice. Interventions designed to promote a supportive learning environment may influence student behavior and academics, educator practice, school climate and discipline practices, and equity concerns such as disproportionate uses of discipline practices. While student outcomes are often the primary focus of the WWC, this review will include outcomes related to the school environment and educator practice, because these outcomes describe supporting learning environments and can influence or mediate student outcomes.

This review focuses on interventions designed to promote supportive learning environments in grades K-12, and will examine the effects of these interventions on schools, staff, and students. Interventions reviewed under this protocol will alter physical, relational, or instructional elements of an educational *environment*. Physical elements of an environment might include security technology or posters that promote appropriate behavior. Relational and instructional elements might include altering discipline policies and instructional strategies. Changes to these elements of a school's environment may improve the perceived sense of trust between students and staff, provide additional emotional support at the school, change disciplinary practices used by school staff, or affect other aspects of school climate.

Of particular interest to this review are interventions designed to improve student behavior or responses to problem behavior, including for students with emotional disturbance or other behavioral risks. To provide WWC users with specific information about an intervention's effectiveness for such students, this review will produce separate reports that present findings for (1) students in this at-risk group and (2) all students.

The following research questions guide this review:

- Which Supportive Learning Environment interventions are effective at improving student outcomes, including engagement in school, academics, behavior, emotional status, and social interactions?
- Which Supportive Learning Environment interventions are effective at improving educators' discipline and instructional practices?
- Which Supportive Learning Environment interventions are effective at improving the school or classroom environment, including measures of school climate and equity with respect to disciplinary actions?
- Are some interventions more effective for certain types of students, or found to be more effective when delivered in certain types of settings?

KEY DEFINITIONS

School environment includes, for purposes of this review, the facilities, physical space, and policies and practices that shape school climate. 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 that might be a target for a Supportive Learning Environment intervention include the physical environment (such as reducing overcrowding, messaging positive behavior expectations on posters), the relational environment (such as improving peer-to-peer communication, or reducing bullying behavior), or the instructional environment (such as improving communication practices or approaches between students and staff).

Non-academic outcomes. Supportive Learning Environment interventions may be designed to improve outcomes outside of academic knowledge and skills. Non-academic outcomes for individuals may span a variety of dimensions, including but not limited to social-emotional learning, perceptions of school climate, bullying, or problem behavior. Other non-academic outcomes might focus on interactions between people and interpersonal relationships (which is sometimes referred to in professional literature as *connectedness*).

Emotional disturbance. This review adopts the definition used in the Individuals With Disabilities Education Act (IDEA), 20 U.S.C. § 1400 (2004) and accompanying federal regulations, 34 C.F.R. § 300.8 (2004). IDEA defines emotional disturbance as a "condition exhibiting one or more of the following characteristics over a long period of time and to a marked degree that adversely affects a student's educational performance:

- (A) An inability to learn that cannot be explained by intellectual, sensory, or health factors.
- (B) An inability to build or maintain satisfactory interpersonal relationships with peers and teachers.
- (C) Inappropriate types of behavior or feelings under normal circumstances.
- (D) A general pervasive mood of unhappiness or depression.
- (E) A tendency to develop physical symptoms or fears associated with personal or school problems.”

States will vary in how they define emotional disturbance. For this review, the followings terms are used to indicate emotional disturbance: emotional or behavioral disorder, serious emotional disturbance, emotional handicap, behavioral disorder, or serious behavior disorder. Students will sometimes be described as being at risk for being classified with emotional disturbance (or similar term) or who have an Individualized Education Plan that specifies a need for an intervention to address a behavior concern. For the purpose of this topic area, these terms/conditions are considered synonymous with emotional disturbance.

ELIGIBILITY CRITERIA

Eligible Populations

Studies that examine the effects of supportive learning environments on general education students and those with specific backgrounds or risk profiles (for example, students with a history of behavioral difficulties) are eligible for the Supportive Learning Environment review. An eligible sample of students may include students with disabilities and/or those receiving special education services. Students must be in grades K-12 (generally ages 5-21) when the supportive learning environment is implemented.

Studies that examine the effects of Supportive Learning Environment interventions on teachers and, more broadly, educators, can also be included in this review. Hence, any teacher or other K-12 school staff member represents 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. In particular, the WWC reports findings from all eligible analyses that meet standards, split into main and supplemental findings. The rating of effectiveness for an intervention is, however,

based on the main findings. Other eligible findings that meet standards can be included in supplemental appendices to the intervention report. For each outcome measure, and among those findings that meet WWC design standards, the 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); and (3) is measured at a time specified by the protocol.

It is anticipated that some studies will provide findings for general education students, and others will focus more on students with emotional disturbance. This review will report on findings where at least 50% of the students are classified as having emotional disturbance, separately from findings for more general populations. In situations where an intervention has evidence for both populations of interest, two intervention reports will be created, where different findings will be highlighted as main findings (instead of supplemental findings) for that given population. For example, if the studies of an intervention present evidence of an intervention's effectiveness for both a whole-school or whole-class sample and for a sample of students with emotional disturbance, the WWC will produce two intervention reports, and each of these results will be shown as a main finding within one report. For the intervention report focused on the whole population, findings for this group will be presented as the main findings, and findings for the students with emotional disturbance will be included as supplemental findings. And similarly, for an intervention report focused on students with emotional disturbance, findings for students with emotional disturbance will be the main findings, and the findings for the whole population will be included as supplemental findings.

Under this review, findings for the following potential subgroups of interest are also eligible to be reported in supplemental appendices in an intervention report. Findings for other subgroups are not eligible for review (unless designated as the main finding based on the criteria above).

Table 1. Subgroups of Interest to the Supportive Learning Environments Topic Area

Characteristics of students	Characteristics of setting or context	Characteristics of educators
Age or grade level Gender Economically disadvantaged (for example, free or reduced-price lunch eligibility, poverty status, or family background) Race or ethnicity English learner status Special education status	Location of the schools involved (for example, urban, suburban, rural) School governance (for example, traditional public, charter, private, religious) Economically disadvantaged school (for example, Title I status)	Position (for example, paraprofessional, teacher, administrator) Years of experience

For this review, measures obtained at the end of an intervention, as well as any time 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. This is because these findings are more prevalent in the studies reviewed under this topic area. 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 also provide useful information about the intervention’s effectiveness. Therefore, follow-up and intermediate findings, when available and appropriate, may be reported in supplemental appendices to the intervention report.

While the above rules will guide how main and supplemental findings are identified, 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

In order to be eligible for review under this topic area, the intervention must in some way seek to change the learning environment. For example, interventions designed to improve school climate, the physical and/or relational environment, reduce conflicts (such as bullying prevention), or promote positive student-teacher interactions, are all eligible for review under this topic area. If an intervention does not include components

that are designed to alter some aspect of the learning environment, it is ineligible for review.

As a consequence, interventions with an exclusive focus on academic improvement will be ineligible for review. For example, a teacher professional development intervention solely intended to improve teacher academic instruction will not be eligible for review under the Supportive Learning Environment topic area. If, however, the professional development was also intended to change aspects of the environment (for example, by reconfiguring the classroom, or changing class behavior/discipline practices), then a study of this intervention would potentially be eligible for review.

To be eligible, the intervention must be delivered in a school setting. Eligible interventions may include, for example, community or parental activities, as long as there is also a school-based component.

Only interventions that are replicable (that is, can be reproduced in another setting) are eligible for review. The following characteristics of an intervention must be documented to reliably reproduce the intervention with different participants, in other settings, and at other times:

- 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 or individual), medium/media of delivery (for example, teacher-led instruction or software), and targeted population;
- Intervention duration and intensity; and
- Description of individuals delivering or administering the intervention.

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

- **Products/programs/services.** The review includes products, programs, or services that can be purchased for school or classroom use as a means to develop a supportive learning environment, including training for school staff, curricula, or other materials. For example, the School Climate Improvement Resource Package would be eligible for review.
- **Practices and strategies.** The review includes schoolwide or classroom-level practices intended to promote a supportive learning environment. For example, Positive Behavioral Interventions and Supports is a set of practices/strategies that would be eligible for review.

- **Policies.** The review includes schoolwide or classroom-level policies, such as enacting an anti-bullying policy, intended to promote a supportive learning environment.

An eligible Supportive Learning Environment intervention may include professional development to support staff delivering the intervention. This review will describe in the intervention reports the professional development provided to staff delivering the intervention based on the information reported in the studies.

Both “branded” and “non-branded” interventions will be reviewed. 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.

Eligible Research

The *WWC Procedures Handbook* discusses the types of research reviewed by the 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:

- **Time frame.** For new intervention reports, the study must have been released within the 20 years preceding the year of the review (in 1999 or later for reviews occurring in 2019). For updated intervention reports, the study must have been released since the original intervention report’s literature search start date (for example, if the original report used 1989 as the literature search start date, the updated report will continue using the same date). Studies must be publicly available (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 grades K-12 at the time that they receive intervention, their outcomes can be measured after they graduate from high school.
- **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 Organizations for Economic Co-operation and Development (OECD) member countries in which English is the primary or most commonly used language (that is, Australia, Canada, Ireland, New Zealand, or the United Kingdom).

Eligible Outcomes

This review includes outcome measures spanning multiple outcome domains. 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.

There are nine outcome domains specified for this Supportive Learning Environment review, organized within three categories: (1) Student Outcomes, (2) Educator Outcomes, and (3) Environmental Outcomes. Table 2 lists the Supportive Learning Environment 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 intervention reports.

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

Student Outcomes	Educator Outcomes	Environmental Outcomes
Student Social Interaction	Educator Discipline Practice	School Climate
Observed Individual Behavior	Educator Instructional Practice	School Equity
Student Emotional Status		
General Academic Achievement		
Student Engagement in School		

Some outcome domains include eligible measures that might reflect perspectives of different stakeholder groups (for example, teachers may report a favorable school climate or that they use fair and equitable discipline practices, but students do not agree). When findings about an intervention’s effectiveness differ by outcome measure within a domain, the WWC may assign the domain the intervention effectiveness rating of *Mixed Effects* if the findings meet criteria indicated in the *WWC Procedures Handbook*. However, regardless of the effectiveness rating given to the domain, the intervention report will describe the pattern of differences in findings within the domain in the written summary of the findings.

Measures based on teachers' observations or self-reports 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. As an example, the Internalizing Problems Subscale of the Behavior Assessment Scale for Children (BASC) 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.

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

Student Outcomes

Student Social Interaction. Includes student behaviors and skills that primarily involve interactions with others, such as students or teachers, or reflect attempts, as well as failed attempts, at social interactions. Examples include:

- observed or perceived peer rejection,
- isolation,
- specific social skills, such as social awareness of context and others, and interpersonal relationship skills, and
- other measures of social behavior that are intended to either benefit (sometimes called prosocial behavior) or harm/hurt others.

Observed Individual Behavior. Includes observed or recordable student behaviors that primarily reflect individual choices and have individual consequences (positive or negative) for the student (sometimes referred to as externalizing behaviors). A key differentiating factor between outcomes in this domain and outcomes in the Student Social Interaction domain are that outcomes in this domain do not necessarily require interactions with others or a social component. Examples of eligible outcomes for this domain include:

- disruptive behavior, including talking out loud at inappropriate times during class,
- impulsivity,
- adaptive functioning and degree of self-control/self-regulation/self-management,
- delinquent behaviors (such as lying, cheating, or stealing),
- physical aggression such as fighting,
- actions characterized as bullying (including physical, relational, cyber),

- sexual harassment,
- arrests,
- substance abuse,
- school suspension or expulsion, and
- following school rules.

Eligible behaviors in this domain can occur in or away from school.

Student Emotional Status. Includes student behaviors and self-ratings that are primarily focused inward and reflect a student’s emotional state (sometimes referred to as emotional or internalizing behaviors), including both positive and negative feelings. Emotional status 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 of emotional status measures include:

- self-awareness of thoughts, feelings and behavior,
- thought disorders,
- emotional regulation,
- depression,
- anxiety,
- happiness, and
- overall adjustment/well-being.

General Academic Achievement. Includes any academic measure based on student test scores from any subject (including mathematics, reading, science, and social studies) or across multiple subject areas (such as the Terra Nova total score).

Student Engagement in School. Student engagement in school is demonstrated through behaviors that are observed during time in school and often reflect school connectedness to academic or extracurricular activities. Examples include:

- school attendance,
- truancy or absences,
- coming to class prepared,
- paying attention in class,
- staying on task during a class assignment,
- completing assignments (including homework), and

- participating in classroom and other school activities.

Engagement is also demonstrated when students indicate that they put effort into being successful in school.

May 2020 Revisions to Student Outcome Domains

The WWC released a Supportive Learning Environment review protocol to be used in conjunction with the *WWC Procedures and Standards Handbooks (version 4.0)* in May 2020. This updated version of the protocol includes revisions to clarify how some student behavior outcomes should be classified into domains. In particular, the descriptions of three domains—student social interaction, observed individual behavior, and student engagement in school—in the previous version of the protocol unintentionally allowed the same behavior to be eligible under more than one outcome domain. The current version of the protocol includes revised descriptions of these three student outcome domains. For example, disruptive behavior is now included as an example of an outcome in the observed individual behavior domain, regardless of when and where the behavior takes place.

Educator Outcomes

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 number of office discipline referrals used by a teacher,
- Relevant subscales about discipline practice from measures based on teacher self-reports, classroom observations, or student surveys.

Educator Instructional Practice. Includes 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 based on surveys administered to students. Eligible assessments of the quality of teacher instruction include, but are not limited to:

- 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 use of the practices 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 of the 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).

Environment Outcomes

School Climate. Includes measures that describe the environment of the school or classroom. Eligible measures in this domain include observations or assessments of the school environment, the quality of interpersonal relationships within the school, 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 school climate, for this review measures of equity will instead potentially be eligible under the school equity domain.

School Equity. 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 office discipline referrals or suspensions differs between two groups of students within a school) and student or staff surveys of their perceptions of equity.

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 the 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 *cautious* boundary for attrition. This boundary was based on the assumption that attrition in studies of Supportive Learning Environment interventions is likely due to factors strongly related to intervention status. Above and beyond factors like family mobility or student absences on days that assessments are conducted, we expect that Supportive Learning Environment interventions may influence whether students or staff remain in schools. For example, victims of bullying may be more inclined to transfer out of schools, and Supportive Learning Environment interventions may affect the prevalence of bullying behavior. Similarly, teachers may be more inclined to leave schools with poor climate or with student behavior problems. As such, it is likely that Supportive Learning Environment interventions will affect the likelihood that students and teachers are respondents in follow-up assessments, and therefore, the *cautious* boundary is appropriate for this topic area.

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 (RCTs)

The 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 topic area, 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, the WWC sometimes differentiates between *early joiners* and *late joiners*. For this review protocol, students will be considered as *early joiners* if they enter a cluster within a 6-week timeframe after the results of random assignment are known, or, in cases where 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 RCTs where the unit of assignment is smaller than a school, *all joiners in the analytic sample pose a risk of bias*. This is 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 where 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 topic area are described below.

- If the intervention is unlikely to directly affect enrollment or placement decisions, then *only late joiners pose a risk of bias*. Joiners who enter clusters after 6 weeks will be considered as 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 includes 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 the Supportive Learning Environment topic area, 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 the 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, the review team leadership has discretion to make a decision on 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:

- 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 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?”, respectively.
- Section 5 of the WWC Standards for reviewing complier average causal effect 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

1. 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 pre-intervention (or baseline) characteristics:

- a) A pre-intervention measure of the outcome used in the analysis, when available. This approach should be used first by reviewers.
- b) A pre-intervention measure of a different variable from within the same domain as the outcome used in the analysis. For example, a pre-intervention measure of peer rejection could be used to show baseline equivalence of an analysis of isolation, since both peer rejection and isolation are found in the Student Social Interaction domain. If the previous approach is not possible, reviewers should use this approach.
- c) Two pre-intervention measures from within two other distinct student outcome domains. This third approach is available only for outcome measures that fall in one of the five domains within the Student Outcomes category (see the list in Table 2). For example, one pre-intervention measure in the *General Academic Achievement* domain and a second in the *Student Engagement in School* domain can be used to satisfy the baseline equivalence requirement for outcomes 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, the WWC will also consider the size of baseline differences for the same analytic sample and from the same domain as the outcome, or from other domains for student outcomes. 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 pre-intervention measure in a given domain, all findings in that domain for that analytic sample will be rated *Does Not Meet WWC Group Design Standards*. Second, for student outcomes, the WWC will also measure the baseline difference between intervention and comparison groups for each eligible pre-intervention measure in a student outcome domain. If the baseline effect sizes are greater than 0.25 standard deviations for any two pre-intervention 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* pre-intervention 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 pre-intervention difference for B requires statistical adjustment, then the impact analyses for A, B, and C must all adjust 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 pre-intervention measures from two other student outcome domains), the analysis must include statistical adjustments for *any* pre-intervention measures within those two student outcome domains with baseline differences in the adjustment range.

In addition to the pre-intervention measures that are required for satisfying the baseline equivalence requirement, a large baseline difference on some other characteristics could be evidence that the intervention and comparison groups are not sufficiently comparable for the purposes of the review. For the Supportive Learning Environment topic area, these characteristics include measures of students' age, grade level, and emotional disturbance status. When differences in student age, grade level, or emotional disturbance status 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, the review team leadership has the discretion to conduct an author query to obtain information on the similarity of the groups on age and grade level.

2. 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 Supportive Learning Environment topic area, 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):

- a) The analytic sample of individuals from any pre-intervention time period.
- b) 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 year prior to when clusters were assigned to conditions.
- c) 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 latest period before the start of the intervention. If authors provide baseline information for multiple samples (e.g., the analytic sample and an adjacent cohort), a reviewer should assess baseline equivalence using the sample listed first in the list 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 review team leadership 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 office discipline referrals or attendance) will generally be considered to meet reliability requirements.

Statistical Adjustments

The *WWC Procedures Handbook* discusses the types of adjustments made by the 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), this topic area uses 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, QEDs), RDDs, or single-case study designs are eligible for review using the appropriate WWC design standards.

PROCEDURES FOR CONDUCTING THE LITERATURE SEARCH

The *WWC Procedures Handbook, Version 4.0*, discusses the procedures for conducting a literature search in Section III: Identifying Relevant Literature and Appendix B: Policies for Searching Studies for Review. This review will use a quick literature search process to identify research on a limited number of interventions that may be of most interest to decision makers, rather than using a broad keyword search on the full topic area to identify interventions. In the first step of this process, content experts identify and recommend interventions with a large body of causal evidence likely to be of interest to decision makers. This review will identify additional interventions that may be the focus of WWC-reviewed studies that are not already the subject of up-to-date WWC intervention reports.

After identifying these interventions, the second step of the process is to conduct intervention-specific literature searches, using the intervention name, to identify all publications on each intervention. This review may refine the potential scope of this search by including additional search terms. For example, terms applicable for the Supportive Learning Environment area include phrases including, but not limited to “school safety, school climate, anti-bullying, or non-academic outcomes.”

In a third step, each citation gathered through this search process undergoes 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 interventions are prioritized for review based on the quantity and quality of eligible studies of the intervention. This prioritization process is described in Appendix A of the *WWC Procedures Handbook*.

Additional Sources

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

- Academy for Social-Emotional Learning in Schools
- American Academy of Pediatrics
- American Educational Research Association
- American Psychological Association

- The School Superintendent Association
- ASCD: Professional Learning & Community for Educators
- Association for Positive Behavior Support
- Centers for Disease Control and Prevention
- Children’s Hospital of Philadelphia
- Channing Bete
- Collaborative for the Advancement of Social and Emotional Learning (CASEL)
- Committee for Children
- Council for Exceptional Children
- Edutopia
- Hazeldon
- 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 School Climate Center
- Positive Behavior Intervention Supports
- School-Counselor.org
- Stopbullying
- The School Superintendent Association

References

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/>.

The National Center on Safe Supportive Learning Environments (n.d.) Retrieved from <https://safesupportivelearning.ed.gov/topic-research/environment>.