

REVIEW PROTOCOL FOR PRIMARY MATHEMATICS VERSION 4.0 (DECEMBER 2018)

This protocol guides the review of research that informs the What Works Clearinghouse (WWC) Primary Mathematics (PM) intervention reports. The protocol is used in conjunction with the [*WWC Procedures and Standards Handbooks \(version 4.0\)*](#).

PURPOSE STATEMENT

Large numbers of students are not proficient in mathematics and there are disparities in mathematics achievement between students from different ethnic and socioeconomic groups. Mathematics skills are important for both academic and workplace success and there are a large number of interventions that aim to improve student mathematics achievement. This review focuses on mathematics interventions that are typically used in kindergarten through grade 8 and are designed to affect students' mathematics achievement.

Whether an intervention falls under the scope of this review depends on both the grade level and content of the intervention. For example, the review can include students in grades 9–12 if they use a primary mathematics intervention such as pre-algebra.

The following research questions guide this review:

- Which interventions are effective at increasing the learning of primary mathematics content and skills?
- Are some primary mathematics interventions more effective for certain types of students, particularly students who are at risk of failure in mathematics?

KEY DEFINITIONS

Mathematics content area. In this review, a mathematics content area is defined as the mathematics topics that are the foundation of a mathematics course, such as number sense, measurement/data, algebra, geometry, trigonometry, or calculus.

Mathematics intervention. In this review, a mathematics intervention is defined as a replicable instructional program that delivers primary mathematics content, clearly delineates mathematics learning goals for students, and is designed to affect student mathematics achievement.

Primary students. Primary students are defined as students in classes where mathematics is presented through multi-topic materials and curricula. In most cases, these courses are taught to students in kindergarten through grade 8, although students in higher grades (such as grades 9 and 10) are considered primary students if they were in multi-topic mathematics courses that cover primary mathematics content, such as pre-algebra.

Secondary students. Secondary students are defined as students in courses that are organized by mathematics content area (for example, algebra, geometry, trigonometry, calculus). In most cases, these courses are taught in grades 9–12, although students in lower grades (such as grades 7 or 8) are considered secondary if they were in a course organized by secondary mathematics content area. Secondary students also include students in integrated high school mathematics courses that include algebra and other secondary content areas.

ELIGIBILITY CRITERIA

Eligible Populations

Studies that examine students who are on track to develop grade- or age-level math skills, who are above grade- or age-level skills, who are behind grade- or age-level skills, who include English Learners (ELs), and who include students with learning disabilities are eligible for the Primary Mathematics review. In this review, the following populations are of interest:

- **Location.** The intervention must be provided to students in a formal or informal school setting, summer school program, or home-school program.
- **Grade range.** The Primary Mathematics area will review studies of interventions designed to increase the learning of primary mathematics content (for example, first-grade mathematics, fourth-grade mathematics, pre-algebra) and skills. In most cases, these interventions are used in kindergarten through grade 8, although students in higher grades (such as grades 9 and 10) fall within this review if they were in multi-topic mathematics course (such as pre-algebra) that do not include secondary mathematics content areas.
- **Overlap between the Primary and Secondary Mathematics topic areas.** Studies of mathematics interventions administered to students in primary grades fall within the scope of the topic area reviews for Secondary Mathematics if the intervention is focused on secondary mathematics content (for example, algebra, geometry, trigonometry, and calculus). Studies that examine the average effect of two or more mathematics interventions that span the Primary and Secondary Mathematics topic areas (such as studies where a portion of the students in the intervention group used pre-algebra and the remaining portion used algebra) will not be reviewed unless disaggregated results can be obtained by intervention course (or mathematics content area). However, longitudinal studies that examined the cumulative effect of two or more mathematics interventions that span the Primary and Secondary Mathematics topic areas will be reviewed based on the intervention course (or mathematics content area) in which the posttest was administered. For example, studies where students used pre-algebra in one grade and algebra in another, and that only posttested students after the algebra course, will be reviewed under Secondary Mathematics.
- **Overlap with other topic areas.** Studies that include a majority of students classified as having a disability and receiving special education services, or a majority of students classified as ELs can be eligible for review under this review protocol. However, review team leadership may determine that the study is ineligible for

review if the intervention of interest was modified, or if the educational setting is focused solely on providing instruction to students with disabilities or who are ELs.

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 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 one finding or set of findings as the main finding: (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.

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

Potential subgroups of interest for this review include:

Characteristics of students	Characteristics of setting or context
<ul style="list-style-type: none"> • Baseline mathematics achievement • Grade • Gender • Socioeconomic status • Race/ethnicity • Special student designation (such as English learner, special education, or at risk in mathematics achievement) 	<ul style="list-style-type: none"> • Location of the schools involved (for example, urban, suburban, rural) • Homogenous groupings of students (for example, tracking) • School type (public, private, religious) • School SES (for example, Title I school) • Average school or class size (small, medium, large) • Average teacher characteristics (for example, teacher education and experience)

Intermediate findings based on eligible measures are available after the start of the intervention are admissible for review. When reported, this review will classify immediate post-intervention findings (for example, outcomes administered after the third year of a 3-year intervention is completed) as main findings because these findings are most prevalent in the studies reviewed under this topic area. Measures occurring several months or years 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

Only interventions that are replicable are eligible for review. The following characteristics of an intervention must be known 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 (for example, whole group, individual), medium/media of delivery (for example, teacher-led instruction or software), and targeted population
- Intervention duration and intensity
- Qualifications of individuals delivering or administering the intervention (for example, teachers or para-professionals)

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

- ***Products and programs (including curricula).*** The review includes curricula or products such as textbooks, software programs, or other educational technology that is (1) intended as the primary mathematics instruction program or (2) designed to supplement the primary classroom program with differentiated instruction, remediation, or enrichment. Examples include *DreamBox Learning*; *Everyday Math*; and *Accelerated Math*.
- ***Practices, strategies, or policies.*** The review includes both general and targeted practices, strategies, and policies. For example, a general practice could be used with a wide range of students and to address a wide range of learning goals. A targeted practice is intended to support instruction for a particular type of student or a particular learning goal for a narrowly defined knowledge or skill. Both general and targeted practices, strategies, and policies must be clearly described and commonly understood in the field and in the literature. Examples of primary mathematics practices include *explicit instruction* and *Peer Assisted Learning Strategies (PALS)*.

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. In this review, the following additional parameters define the scope of research studies to be included:

- **Topic.** The study must focus on the effects of a mathematics intervention on one or more measures of mathematics achievement.
- **Time frame.** For new intervention reports, the study must have been released within the 20 years preceding the year of the review (for example, 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 a 1989 literature search start date, the updated report will continue using the same date). Studies must be publicly available (accessible online 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 described in the “Eligible Populations” section at the time they receive the intervention. Outcomes can be measured later (for example, when the sample is older).
- **Language.** The study must be available in English to be included in the review. Studies examining mathematics competencies in other languages will 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 a country that is sufficiently similar to the United States that the study could be replicated in the United States.

Eligible Outcomes

This review includes outcomes in the following domains:

Domain name	Description
<i>Number and operations</i>	Includes understanding numbers and integers, such as subitizing, estimation, number order, number combinations, counting, comparisons, operations, computing fluently, representing fractions and ratios and understanding the base-ten number system and fractions.
<i>Geometry and measurement</i>	Includes two-dimensional and three-dimensional geometric shapes and understanding properties, composition, and geometric relationships, including visualization, spatial reasoning, and geometric modeling. Also includes understanding the attributes, units, systems, and processes of measurement, and applying techniques, tools, and formulas to determine measurements.

Domain name	Description
<i>Data analysis, statistics, and probability</i>	Includes collecting, organizing, and displaying data to answer questions. Also includes statistical methods to analyze data, making inferences and predictions based on data, and the basic concepts of probability.
<i>Algebra</i>	Includes patterns or functions, such as identifying, representing, or applying a function or “rule”. Although studies of algebra outcomes measured after an algebra course are not eligible for review under this protocol (because they are eligible for review under the secondary mathematics protocol), studies of primary mathematics interventions may include outcomes that measure algebraic concepts, and are eligible for this review.
<i>General mathematics achievement</i>	Outcomes that fall in the mathematics achievement domain combine separate measures of two or more of the previous content domains (number and operations; measurement; geometry; or data analysis, statistics, and probability) by providing a summary score across domains, such as a “total math score” on a standardized mathematics test. This domain also includes tests of mathematical understanding, procedures, and problem solving that are designed to measure more than one content area.

Relevant outcome measures of student mathematics achievement in the above domains include standardized, nationally-normed achievement tests that are appropriate for students in kindergarten through grade 8; standardized state or local tests of mathematics achievement; and research-based or locally-developed tests or instruments that assess students’ mathematics concepts or skills in the above domains.

Other measures of mathematics performance, such as student grades assigned by teachers, are not eligible for review.

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 (RDD) in Section III.C

This review uses the *optimistic* boundary for attrition. This boundary was selected based on the assumption that most attrition in studies of Primary Mathematics was due to factors that were not strongly related to intervention status. For example, these factors may include family mobility or absences on days that assessments are conducted. 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. Based on the choice of the boundary, the study review guide calculates attrition and whether it is high or low.

Joiners in Cluster Randomized Controlled Trials (RCTs)

The WWC defines a *joiner* as any student who enters a cluster (for example, a school or a classroom) after the results of random assignment are known to any individual who could plausibly 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.

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, we will consider a student to be an *early joiner* if they enter a cluster in the 6 weeks 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. That is, the *early period for joiners ends 6 weeks after the start of the school year if the results of random assignment were announced over the summer; otherwise, the early period ends 6 weeks after the results of random assignment were announced*. *Late joiners* are those that enter clusters after the end of the early period.

This review protocol specifies the following rules:

- a. In cluster RCTs where the unit of assignment is a classroom or another group defined within a school (such as a group of classrooms or a small group of students within classrooms), ***all joiners 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 their assignment to 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 does not limit the risk of bias from joiners.
- b. In cluster RCTs where the unit of assignment is a school or a group of schools (such as a district), whether joiners pose a risk of bias depends on whether the intervention is expected

to influence school enrollment or placement decisions. Two most common examples for this topic area are discussed below.

- If the intervention may affect enrollment or placement decisions (such as a magnet program with a mathematics focus), then ***all joiners pose a risk of bias***. A study of such an intervention that includes one or more joiners in the analytic sample ***does not limit the risk of bias from joiners***.
- If the intervention is a curriculum or another intervention used in all classrooms (such as *Saxon Math*[®]) that does *not* directly affect enrollment or placement decisions, then ***only late joiners pose a risk of bias***. Late joiners may be more likely to do so because of the intervention, and therefore differ from those who join the comparison group. A study of such an intervention that includes at least one late joiner in the analytic sample ***does not limit the risk of bias from joiners***.

For Primary Mathematics reviews, the default assumption is that the interventions being examined with assignment at the school-level or higher are unlikely to affect enrollment or placement decisions; however, review team leadership has discretion to revise this assessment.

Additionally, typical scenarios the WWC encounters in cluster RCTs are described above, but we cannot anticipate all scenarios. 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 (QED), the study must demonstrate baseline equivalence of the intervention and comparison groups for the analytic sample. The onus for demonstrating equivalence in these studies rests with the authors. The *WWC Standards Handbook* discusses how authors must demonstrate baseline equivalence 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 RDDs 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 on one of the following pre-intervention (or baseline) characteristics:

- A pre-intervention measure of the outcome used in the analysis; or
- If a pre-intervention measure of the outcome used in the analysis is not available, a pre-intervention measure of an outcome from any of the five outcome domains detailed in the “Eligible Outcomes” section above can be used. For example, a pretest from the general mathematics achievement domain can be used to establish baseline equivalence for a number and operations outcome when a number and operations pretest is unavailable.

This review assesses baseline equivalence within each analytic sample and for each individual outcome measure. In particular:

- Because it is common for studies in this topic area to assess the same or highly related outcome measure at baseline and follow-up, this review assesses baseline on an outcome-by-outcome basis. Specifically, baseline equivalence for an eligible outcome measure is assessed based on the magnitude of baseline differences for its most closely associated pre-intervention measure. Baseline differences on other pre-intervention measures do not influence the assessment of baseline equivalence for the outcome measure. For example, if both pre- and post-intervention measures of outcomes A, B, and C are available and the baseline difference for the pre-intervention measure of C is outside of the statistical adjustment range (that is, it exceeds 0.25 standard deviations), then the finding for this outcome would be rated *Does Not Meet WWC Design Standards* because it does not satisfy the baseline equivalence requirement. A finding for outcome B, however, could still meet standards if the authors satisfy the baseline equivalence requirement using the pre-intervention measure of B.
- Similarly, this review requires that, when the baseline difference for a pre-intervention measure is in the statistical adjustment range (that is, it is between 0.05 and 0.25 standard deviations), the adjustment must be made only in the analysis of the associated outcome measure. For example, if the pre-intervention difference in B requires statistical adjustment, only the analysis of outcome B must adjust for B.

In addition to the pre-intervention measures that are required for satisfying the baseline equivalence requirement, other sample characteristics such as student age and grade level, may be associated with the outcome. A large baseline difference on these characteristics could be evidence that the intervention and comparison groups are not sufficiently comparable for the purposes of the review. When differences in student age or grade level are larger than 0.25 standard deviations, the study will be rated *Does Not Meet WWC 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 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 samples for demonstrating baseline equivalence of clusters

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 in the clusters at the time the baseline data were collected):

- a. The analytic sample of individuals from any pre-intervention 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, 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

The *WWC Standards Handbook* discusses the types of outcomes, criteria the outcome must meet, and how outcomes are reported in Section IV.A: Outcome Requirements and Reporting. In particular, this review follows the requirements stated in the *Standards Handbook* regarding the reliability of outcome measures.

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 WWC default intra-class correlation coefficient for achievement outcomes of 0.20 for all eligible outcomes unless a study-reported intra-class correlation coefficient is available.

Eligible Study Designs

Studies that use group designs (RCTs and QEDs), RDDs, or single-case designs (SCDs) are eligible for review using the appropriate standards or pilot 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, such as the word ‘mathematics,’ the grade levels targeted by the intervention, and terms to reference the nature of the intervention (such as curriculum, supplemental program, or instructional practice).

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 websites and electronic databases listed in Appendix B of the *WWC Procedures Handbook* as well as the following websites:

- American Association of Colleges of Teacher Education (AACTE)
- American Association of School Administrators (AASA)
- American Federation of Teachers

- American Mathematical Society (AMS)
- American Statistical Association (ASA)
- Association for Supervision and Curriculum Development (ASCD)
- Broad Foundation (Education)
- Center for Research in Educational Policy (CREP)
- Center for the Study of Instructional Improvement
- Center on Education Policy
- Center on Instruction
- Consortium for Policy Research in Education (CPRE)
- Council of Chief State School Officers
- Council of the Great City Schools (CGCS)
- Editorial Projects in Education (EPE) Research Center
- Education Development Center (EDC)
- Erikson Institute, University of Chicago
- Geometry, Reasoning, and Instructional Practices
- Harvard Graduate School of Education
- Johns Hopkins University School of Education
- Mathematical Association of America (MAA)
- Mathematics Teaching and Learning to Teach Project
- Mid-continent Research for Education and Learning
- Millennium Mathematics Project
- National Association for the Education of Young Children
- National Association of Elementary School Principals (NAESP)
- National Association of Secondary School Principals (NASSP)
- National Board for Professional Teaching Standards
- National Center for Children in Poverty
- National Center for Education Research
- National Center for Research on Early Childhood Education (NCRECE)
- National Center for Special Education Research
- National Conference of State Legislatures (NCSL)
- National Council of Supervisors of Mathematics (NCSM)

- National Council of Teachers of Mathematics (NCTM)
- National Head Start Association
- National Math Panel
- National Science Foundation (NSF)
- New America Foundation's Early Education Initiative
- Office of Early Learning (<http://www.ed.gov/early-learning>)
- Pacific Resources for Education and Learning (PREL)
- Promising Practices Network
- Public Education Network
- Public Policy Research Institute at Texas A&M University
- Society for Research in Child Development
- Southwest Educational Development Laboratory (SEDL)
- Technical Assistance Center on Social Emotional Intervention for Young Children
- TERC
- The National Academies Press
- U.S. Department of Education
- U.S. Department of Health & Human Services
- UCLA Graduate School of Education Research Centers
- University of Chicago School Mathematics Project
- WestEd