REVIEW PROTOCOL FOR PRIMARY SCIENCE
VERSION 4.0 (MARCH 2019)

This protocol guides the review of research that informs the What Works Clearinghouse (WWC) Primary Science (PS) intervention reports. The protocol is used in conjunction with the *WWC Procedures Handbook (version 4.0)* and the *WWC Standards Handbook (version 4.0).*

PURPOSE STATEMENT

Large numbers of students lack proficiency in science, and students from different ethnic and socioeconomic groups show disparities in science achievement. Science knowledge and skills are important for both academic and workplace success, and a number of interventions aim to improve student science achievement. This review focuses on science interventions used in kindergarten through grade 8 and designed to affect students' science achievement.

The following research questions guide this review:

- Which interventions are effective at increasing the learning of science content, processes, and skills for students in kindergarten through grade 8?
- Are some primary science interventions more effective for certain types of students, particularly those who are members of traditionally underrepresented groups in science?

KEY DEFINITIONS

Authentic educational settings. These are locations—including schools and alternative schools, school systems, and settings—that deliver supplemental or remedial education services during *regular school hours*.

Out-of-school-time settings. These are activities or programming that occur in non-school *locations*, at non-school *times*, or both. Out-of-school-time settings include museums, public libraries, community or recreational centers, university or college campuses, or academic or industry research laboratories. Non-school times are periods outside of regular school hours (for example, after school, on weekends, or during summer). An afterschool program that takes place within the school building or other authentic education setting is defined as an out-of-school-time activity for purposes of this review.

Primary students. Primary students are those in grades kindergarten through grade 8. Within each of these grades, the science curriculum and instruction typically cover some combination of topics from the physical, life, and earth/space sciences.

Secondary students. Secondary students are defined as students in grades 9 to 12 who are enrolled in science courses organized by major scientific discipline (for example, physics, biology, or chemistry); in some schools, content from more than one discipline is combined (for example, physics and chemistry or biochemistry) or may have prerequisites (for example, forensic science).

Scientific content area. In this review, science refers primarily to the physical, life, and earth and space sciences and also includes topics such as the nature of scientific inquiry (for example, control of variables, analyzing data, constructing explanations, and arguing from evidence). Science content areas are topics that are fundamental to understanding:

- The nature of scientific inquiry (science "practices")
- Physical systems (forces, motion, and stability; energy; wave forms)
- The structure of matter (substances and their properties, constituents of matter, interactions of matter with energy)
- Biological systems (structures, functions and processes of living things from molecules to organisms; variation and inheritance of traits; energy and dynamics in ecosystems; evolution and the principles of unity and diversity)
- Earth and space systems (weather and climate; physical geography and geology; Earth and human activity; Earth and its place in the solar system and the universe).

Science intervention. In this review, a science intervention is defined as a replicable instructional program, product, practice, strategy, or policy that clearly delineates science learning goals for students and is designed to affect student science achievement.

ELIGIBILITY CRITERIA

Eligible populations

Studies eligible for the Primary Science review are those that examine students who are on track to develop grade- or age-level science achievement, who demonstrate above grade- or age-level science achievement, who are behind grade- or age-level science achievement, who include English learners, and who include students with learning disabilities. In this review, the following populations are of interest:

- ***Location.*** The intervention must be provided to students in an authentic educational setting or out-of-school-time setting.
- ***Grade range.*** The Primary Science review will include studies of interventions designed to increase the learning of science content and practices in kindergarten through grade 8.
- ***Overlap between the Primary and Secondary Science topic areas.*** Studies of science interventions administered to students in primary grades fall within the scope of topic area reviews for Secondary Science if the intervention is focused on a secondary science content area. Studies that examine the average effect of two or more science interventions that span Primary and Secondary Science topic areas will not be reviewed unless disaggregated results can be obtained for each intervention course or topic area. However, longitudinal studies that examine the cumulative effect of two or more science interventions spanning the Primary and Secondary Science topic areas will be reviewed based on the intervention course (or science content area) in which the post-test was administered. For example, studies where students received 8th grade-level science

instruction in 8th grade and 9th grade biology instruction in 9th grade and that only post-tested students after the biology course will be reviewed under Secondary Science.

- ***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 English learners 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 English learners.

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, from multiple analyses that use composite or subscale scores, or from 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 three criteria to designate a finding or set of findings as the main finding: it (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).

Exhibit 1. Subgroups of Interest to the Primary Science Topic Area

Characteristics of students	Characteristics of setting or context
• Baseline science achievement • Grade • Gender • Socioeconomic status • Race/ethnicity, especially: • Hispanic or Latino • Black or African American • Native Hawaiian or other Pacific Islander • American Indian or Alaska Native • Special student designation (such as English learner, special education, or at-risk in science achievement)	• Location of the schools or setting (for example, urban, suburban, rural) or characteristics of out-of-school-time host/sponsoring institutions; district or community-based sponsor such as museum, library, or local university) • Timing of participation (for example, regular school hours, after-school, weekends, summer programs) • Homogenous groupings of students (for example, tracking) • School SES (for example, Title I school) • Average teacher characteristics (for example, teacher education and experience)

Intermediate findings based on eligible measures available after the start of the intervention are admissible for review. When reported, this review will classify findings for outcomes administered immediately after the intervention (for example, outcomes administered after completion of the third year of a three-year intervention) as main findings because these findings are most prevalent in the studies reviewed under this topic area. Follow-up measures administered 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 the 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: skill being targeted, approach to enhancing the skill (for example, strategies, activities, and materials), unit of delivery of the intervention (for example, whole group or 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 paraprofessionals).

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 intended as the primary science instruction program or designed to supplement the classroom material with differentiated instruction, remediation, or enrichment. Examples of science curricula include *HMH Science Dimension; Full Option Science System;* and *Elevate Science*. An example supplemental product is *Science Laboratory, SRA: Life, Earth, Physical*.
- ***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 science practices include inquiry-based instruction, project- or problem-based learning, and laboratory work.

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

Some interventions, especially those targeting students in kindergarten through grade 8, may integrate science content with content in mathematics, technology, or engineering. Studies of such interventions are eligible for review under the Primary Science protocol so long as learning goals are articulated for one or more scientific domains and the interventions are designed to affect student achievement in science.

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

- ***Topic.*** The study must focus on the effects of a primary science intervention on one or more measures of science 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) and be obtained by the WWC for review prior to the drafting of the intervention report. 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 include an eligible population, as defined earlier. Outcomes can be measured after the intervention (for example, when the sample is older).
- ***Language.*** The study must be available in English. However, instruction may be delivered in other languages and studies examining science competencies using measures administered in other languages will be included in the review.
- ***Location.*** The study must include students in the United States, 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 three disciplinary domains and a general science achievement domain for outcomes that combine two or more of the disciplinary domains or cut across disciplinary boundaries. Exhibit 2 describes these domains.

Exhibit 2. Outcome Domains for the Primary Science Topic Area

Domain name	Description
Life sciences	Includes the structures and functions of living things at different scales; growth, development, and reproduction of organisms; information processing and behavior in organisms; matter and energy transfer in living things and ecosystems; inheritance of and variation in traits; natural selection and adaptation; evidence of common ancestry; biodiversity
Physical sciences	Includes the properties of matter and changes in matter; force, motion, and interactions of forces; energy and energy transfer and conservation; relationship between energy and forces; properties of waves; electromagnetic radiation
Earth/space sciences	Includes the structures, properties and materials of Earth; tectonics; Earth's place in the solar system and the universe; changes in Earth over time; water, weather and climate; energy in Earth systems; bio-geology
General science achievement	Outcomes related to science content in two or more of the domains listed earlier as well as concepts that cut across disciplinary boundaries (for example, the role of patterns; cause and effect relationships; stability and change in natural systems); includes knowledge of science practices, such as forming hypotheses and making predictions, control of variables, and planning and conducting systematic investigations (experiments and observations)

Relevant outcome measures of these science domains include standardized, nationally normed achievement tests that are appropriate for students in kindergarten through grade 8; standardized state or local tests of science achievement; and research-based or locally developed tests or instruments that assess students' understanding of science concepts or practices in the above domains. Other measures of science achievement, 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—"Standards for a Single RRD Impact."

Under this review, the boundary for attrition depends on the educational setting in which an intervention takes place:

- For studies of interventions in authentic educational settings, this review uses the optimistic boundary for attrition.
- For studies of interventions in out-of-school-time settings, this review uses the cautious boundary for attrition.

For authentic educational settings, we selected the optimistic boundary based on the assumption that most attrition in studies of primary science in such settings is due to factors that are not strongly related to intervention status. For example, most attrition in studies of primary science interventions results from exogenous factors, such as family mobility or student absence on days that assessments are conducted. In contrast, we selected the cautious boundary for out-of-school-time settings because attrition may be related to intervention status in those settings. For example, compared to interventions implemented by schools or districts during regular instructional hours, for interventions in out-of-school-time settings, parents or students may be more likely to object to the results of random assignment (for example, by leaving the comparison condition). As a result, attrition could result from endogenous factors related to the nature of the intervention.

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 students to be *early joiners* if they enter a cluster in the six weeks after the results of random assignment are known, or, if random assignment occurred during the summer, six weeks after the start of the school year. *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 groups of classrooms or small groups 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. The two most common examples for this topic area are the following:

- If the intervention may affect enrollment or placement decisions (as with a magnet program with a science and technology focus or an afterschool or summer school science program), 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 *FOSS Science curriculum*®) 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 the Primary Science reviews, the default assumption is that interventions in authentic educational settings 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. For interventions in out-of-school-time settings, the default assumption is that the intervention may affect enrollment or placement decisions because of the possibility that parents could find science interventions in out-of-school-time settings particularly attractive (and join the intervention group outside of the random assignment process).

The two examples just reviewed reflect the typical scenarios the WWC encounters in cluster RCTs, but we cannot anticipate all scenarios. When an intervention and unit of assignment in a cluster RCT do not fall into one of these two categories, 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 either an RCT or a 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 study authors. The *WWC Standards Handbook* discusses how authors must demonstrate baseline equivalence in these locations:

- 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—"Continuity of the Relationship Between the Outcome and the Forcing Variable"

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 a pre-intervention (baseline) measure of the outcome used in analysis.

Because outcomes within a disciplinary science domain (life, physical or Earth/space sciences) may be conceptually distinct from outcomes from another disciplinary domain, this review assesses baseline equivalence on a domain-by-domain basis. The following pre-intervention measures may be used:

- A pre-intervention measure of the outcome
- For an outcome in any of the four primary science domains, another pre-intervention measure from any one of the following domains:
  - The same domain as the outcome
  - General science achievement, as defined earlier
  - A standardized measure of reading achievement (but not writing achievement) from the general literacy achievement domain or the reading comprehension construct of

the comprehension domain (but not the vocabulary development construct), as defined in the Adolescent Literacy review protocol (version 4.0)

- General mathematics achievement, as defined in the Primary Mathematics review protocol (version 4.0)

For example, for a life sciences outcome, a pre-intervention measure from the life sciences domain, general science achievement domain, general literacy achievement domain (but only standardized reading assessments), reading comprehension construct, or general mathematics achievement domain can be used to establish baseline equivalence. However, a pre-intervention measure from the physical sciences domain or the Earth/space sciences domain may not be used to establish baseline equivalence when the outcome is from the life sciences domain or the general science achievement domain.

The WWC will measure the baseline effect size difference between intervention and comparison groups for each eligible pre-intervention measure reported in the study. If the analytic sample that must satisfy the baseline equivalence requirement has a baseline effect size greater than 0.25 standard deviations for *any* eligible pre-intervention measure, then all findings for this analytic sample in that domain *do 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* eligible pre-intervention measure, *all outcome measures within that domain must adjust for each pre-intervention measure with a baseline difference that falls within the adjustment range*.

For example, consider a study with pre-intervention measures A, B, and C available for an analytic sample in the physical sciences domain that must satisfy the baseline equivalence requirement when:

- A is a pre-intervention measure within the physical sciences domain
- B is a pre-intervention measure within the life sciences domain
- C is a pre-intervention measure of reading comprehension.

With this analytic sample, findings must examine the effect size of the baseline difference between the intervention and control groups for both A and C, but not B. (Measure B, from the life sciences domain, is not an eligible pre-intervention measure for the physical sciences domain.)

- If the pre-intervention difference for *either* A or C is greater than 0.25, then the findings for analytic samples within the physical sciences domain do *not* meet WWC group design standards.
- If the pre-intervention differences for A and C both require statistical adjustment, then the impact analysis must adjust for *both* A and C to be eligible to meet WWC standards with reservations.
- If the pre-intervention difference for A is less than 0.05, but the pre-intervention difference for C requires statistical adjustment, then the impact analysis must adjust for C to be eligible to meet WWC standards with reservations;

- If the pre-intervention differences for A and C are both less than 0.05, then no statistical adjustment is necessary for findings to be eligible to meet WWC standards with reservations.

In addition to the pre-intervention measures 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 measures could be evidence that the populations were drawn from very different settings and 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 to conduct an author query to obtain information on the similarity of the groups in 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 earlier 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 within 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 (in this case, the baseline data may be obtained at the time that clusters were assigned to conditions; during the year prior to when clusters were assigned to conditions; or, with multi-year interventions, during the year prior to the cohort first beginning to receive the intervention)

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 sample (a), if available, then sample (b), and then sample (c).

If authors provide baseline information for multiple samples across multiple time periods, the reviewer should consult the 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, two 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, the 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 in Appendix B: Policies for Searching Studies for Review. We 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 will be to conduct intervention-specific literature searches, using the intervention name, to identify all publications relevant to each intervention. This review may refine the potential scope of this search by including additional search terms, such as the acronym *STEM*; the word *science*, 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 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*: Screening Studies. Finally, the interventions are prioritized for review based on the quantity and quality of eligible studies of the intervention, applying the procedures in Appendix A of the *WWC Procedures Handbook*: Policies for Prioritizing Studies for Review.

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:

- Achieve, Inc.
- Alliance for Excellent Education
- American Association for the Advancement of Science
- American Association of Colleges of Teacher Education (AACTE)
- American Association of Physics Teachers
- American Association of School Administrators (AASA)
- American Federation of Teachers
- Association for Supervision and Curriculum Development (ASCD)
- Best Evidence Encyclopedia (BEE)
- Biological Science Curriculum Study (BSCS)
- Broad Foundation (Education)
- Center for Data-Driven Reform in Education (CDDRE) at Johns Hopkins University
- Center for Research and Exploration in Space Science and Technology (CRESST)
- Center for Research and Reform in Education (CRRE) at Johns Hopkins University
- Center for Research in Educational Policy (CREP)
- Center for the Study of Instructional Improvement
- Center on Education Policy
- Center on Instruction
- Congressional Research Service
- 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
- For Inspiration and Recognition of Science and Technology (FIRST)
- Harvard Graduate School of Education
- JASON Learning
- Johns Hopkins University School of Education
- LinkEngineering
- Mid-continent Research for Education and Learning
- National Academies Press
- National Association for the Education of Young Children
- National Association for Research in Science Teaching (NARST)
- 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 Head Start Association
- National Science Digital Library (NSDL)
- National Science Foundation (NSF)
- National Science Resources Center (NSRC)
- National Science Teachers Association (NSTA)
- New America Foundation's Early Education Initiative
- Office of Early Learning
- Pacific Resources for Education and Learning (PREL)
- PhysPort
- Project Lead The Way (PLTW)
- Promising Practices Network
- Public Education Network

- Public Policy Research Institute at Texas A&M University
- Science Education for Public Understanding Program (SEPUP)
- Society for Research in Child Development
- Southwest Educational Development Laboratory (SEDL)
- TERC
- UCLA Graduate School of Education Research Centers
- U.S. Department of Education
- U.S. Department of Health & Human Services
- WestEd