

WWC EVIDENCE REVIEW PROTOCOL FOR INTERVENTIONS FOR ENGLISH LEARNERS, VERSION 4.0 (FEBRUARY 2020)

This review-specific protocol guides the review of research that informs the What Works Clearinghouse (WWC) intervention reports in the English learners (EL) topic area. The review-specific protocol is used in conjunction with the *WWC Procedures and Standards Handbooks (version 4.0)*.

PURPOSE STATEMENT

ELs are among the fastest-growing population in U.S. schools and comprise 9 percent of all students in grades PreK–12 nationwide (U.S. Department of Education, 2018). Despite the fact that ELs bring many assets with them to school, including their home language, they consistently perform at lower levels than their non-EL peers and have lower high school graduation rates (National Center for Education Statistics, 2014). This review focuses on interventions designed to address this achievement gap by improving the English language and literacy skills and/or content area achievement in mathematics, science, technology, engineering, or social studies of ELs in grades PreK–12.

The following research question guides this review:

- Which interventions improve English language and literacy skills among EL students in grades PreK–12?
- Which interventions improve content area achievement among EL students in grades PreK–12?

KEY DEFINITIONS

English learners. Current ELs are students with a primary language other than English who have a limited range of speaking, reading, writing, and listening skills in English. This includes students who have been identified and determined by their school as having limited English proficiency at the time of the study. Former ELs are those students who have demonstrated English proficiency as determined by their school, but who have been officially exited from EL status. This protocol includes current ELs and former ELs who have been reclassified within the last 2 academic years.

Terms such as limited English proficient (LEP), English language learner (ELL), non-English speaker (NES), English as a Second Language (ESL), English for Speakers of Other Languages (ESOL), language minority (LM), second language learner (SLL), or dual language learner (DLL) may also describe students who are English learners in some contexts.

English language and literacy skills. These skills include speaking, listening, reading, and writing in English.

Content area skills. For the purposes of this protocol, these include skills in the content areas of mathematics, science, engineering, and social studies.

ELIGIBILITY CRITERIA

Eligible Populations

In this review, the following populations are of interest:

- **EL status.** This WWC review includes students who are current ELs or former ELs who have exited EL status within the last 2 years. The study sample must have a subgroup analysis for ELs, or the sample must include at least 50% ELs.
- **Grade range.** Studies with students in grades PreK–12 (or ages 3–21) when they receive the intervention are eligible for review. Studies that include students who are older than 21 or younger than 3 when they receive the intervention are included as long as the students are in grades PreK–12. If authors do not provide the grade for study students, the review will use the age range of 3–21 to determine if the study is eligible.
- **Location.** The intervention must be provided to students in an academic setting, including preschools, elementary schools, secondary schools, summer school programs, or home-schooling programs.

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 results for an EL subgroup or full study sample if at least 50 percent of the sample is EL; (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 are presented in Exhibit 1.

Exhibit 1. Subgroups of Interest to the English Learners Topic Area

Characteristics of students	Characteristics of setting or context
<ul style="list-style-type: none">• Age or grade level• Gender• Economically disadvantaged (for example, free or reduced-price lunch eligibility, poverty status, or family background)• Pre-intervention English language skills• Race or ethnicity• Special education status	<ul style="list-style-type: none">• Location of the schools involved (for example, urban, suburban, rural)• Economically disadvantaged school (for example, Title I status)

Results for native English speakers are also eligible for review as supplemental findings if they are in a study with an eligible main finding for English learners that meets WWC design standards. However, findings for native English speakers cannot contribute to the rating of effectiveness for an intervention.

Also 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 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 more prevalent in most 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. However, the review team has discretion to classify intermediate or follow-up findings as main findings for all studies of an intervention, when the design or content of an intervention suggests that its effects may more likely to appear some time during or after the intervention.

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 (that 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) (such as strategies, activities, and materials), unit of delivery of the intervention (such as whole group, individual), medium/media of delivery (such as teacher-led instruction or software), and targeted population;

- Intervention duration and intensity; and
- Description of individuals delivering or administering the intervention.

Studies of interventions where at least some of the instruction is conducted in a language other than English are eligible for this topic area review. It is a common practice in schools to sometimes deliver instruction in languages other than English in classrooms with English learners. As discussed below, changing the language of instruction alone can be an eligible intervention under this review protocol. Other eligible interventions may change both the language of instruction and other factors. However, when the intervention is a product, program, practice, or policy that does not include a change in the language of instruction, both conditions must be taught in the same languages and for the same proportions of the day.

To be eligible, a study also must report outcomes in English in one of the eligible outcome domains. Outcome measures in the other language will be reported as supplementary findings. However, the WWC rating for the study will be based on the English outcome measures, and only findings for those outcome measures will contribute to the rating of effectiveness for the intervention.

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

- **Products (including curricula).** The review includes (a) English language, literacy, or content area curricula; (b) English language, literacy or content area textbooks intended for whole-school or whole-classroom use; and (c) software (computer or web-based) applications designed to improve language, literacy, and content area skills and used with an entire class or individual students.
- **Programs.** The review includes (a) supplemental programs that are intended to enhance a whole-school or whole-classroom language, literacy, or content area instruction and (b) programs or remedial curricula aimed at struggling readers and students who read behind their grade level (such as an afterschool tutoring program).
- **Instructional practices or strategies.** The review includes instructional practices or strategies, defined as a named approach to promoting the academic development of students. School staff members implement a practice while interacting with students and materials in their classrooms. The named approach must be described and commonly understood in the field. Instructional practices or strategies, such as sheltered instruction, peer assisted learning, and rich vocabulary instruction, are eligible for review.
- **Policies.** This review includes school-wide or district-wide policies intended to improve English language and literacy or content area outcomes (such as a school-wide literacy initiative).
- **Language of instruction.** Studies in which the language of instruction *is* the intervention are eligible for this review. For example, a study that compares the effectiveness (in terms of eligible student outcomes) of students in one- or two-way immersion programs versus students in English-only programs is eligible for review. In these cases, the intervention and comparison groups must be comparable in all other ways, such that the language of instruction is not confounded with other factors.

An eligible intervention may include professional development to support staff delivering the intervention. This topic area 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:

- **Topic.** The research study must focus on the effects of an English language, literacy or content area intervention on one or more measures of ELs’ English language, literacy, or content areas skills.
- **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 1983 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 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 PreK–12 at the time that they receive intervention, their outcomes can be measured after students 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 Organisation 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 outcomes in the domains and corresponding constructs listed below (see Exhibit 2). Some of these domains and constructs are similar, but not identical to, the domains and constructs in other WWC protocols. For example, this review includes constructs for vocabulary and listening comprehension in the English language proficiency domain, whereas the Adolescent Literacy protocol (version 4.0) includes these constructs in the Comprehension domain. In addition, for this review, outcome measures can be administered in English or in students’ native language, although only outcomes in English will be reported as main findings and contribute to the WWC rating for the study, as stated above. Outcomes in other languages can be reported in supplemental appendices if they fall in eligible domains and meet standards.

Exhibit 2. Outcome Domains and Constructs for the English Learners Topic Area

Domain name	Construct name and description
<i>Alphabetic</i>	<i>Letter identification</i> —knowledge of the names of the letters of the alphabet.
	<i>Phonemic awareness</i> —understanding that the sounds of spoken language (or phonemes) work together to make words, and phonemes can be substituted and rearranged to create different words. Phonemic awareness includes the ability to identify, think about, and work with the individual sounds in spoken words. Phonemic awareness helps students learn how to read and spell by allowing them to combine or blend the separate sounds of a word to say the word (for example, “/c/ /a/ /t/—cat”).
	<i>Phonics</i> ^a —(a) knowing that there is a predictable relationship between phonemes (the sounds in spoken language) and graphemes (the letters used to represent the sounds in written language), (b) associating letters and letter combinations with sounds and blending them into syllables and words, and (c) understanding that this information can be used to read or decode words. Spelling is included as an acceptable phonics outcome. Spelling skills also play an important role within writing; however, because spelling skills are necessarily subordinate to other issues of writing quality, the spelling outcomes are included together with decoding skills within the phonics construct in this review protocol.
	<i>Phonological awareness</i> —knowledge of phonemes and of larger spoken units such as syllables and rhyming words. Tests of phonological awareness might require students to generate words that rhyme, to segment sentences into words, to segment polysyllabic words into syllables, or to delete syllables from words (for example, “what is cowboy without cow?”).

Domain name	Construct name and description
Alphabetics (continued)	Print awareness —knowledge or concepts about print, such as (a) print carries a message; (b) there are conventions of print, such as directionality (left to right, top to bottom), differences between letters and words, distinctions between upper- and lowercase, punctuation; and (c) books have some common characteristics (for example, author, title, front/back).
Reading fluency	Reading text accurately, automatically, and with expression (including appropriate pausing, response to punctuation, and so on), while extracting meaning from it. This domain does not have constructs.
Reading comprehension	Reading comprehension —understanding the meaning of a passage. Reading comprehension depends on various underlying components including decoding (the ability to translate text into speech), knowledge of word meanings, fluency (the ability to read text accurately and automatically), and the ability to understand and interpret spoken language. Reading comprehension outcomes may include tests of students’ comprehension of passages from various content areas. For example, a test assessing students’ ability to read and answer questions about a social studies passage would be an acceptable outcome. However, content area knowledge tests that assess student’s preexisting understanding of the facts, theories, and other content related material are not eligible outcomes. ^a Measures of vocabulary development are excluded from this domain and included in the English Language Proficiency domain described below.
Writing conventions	Using rules of standard English language such as word usage, sentence structure, grammar, and language mechanics including capitalization, punctuation, or spelling. Standardized tests using multiple-choice formats that test a student’s knowledge of writing conventions (such as English language arts or written language subtests) may fall within this domain. When spelling skills are assessed based on students’ writing samples, they are included in the writing conventions domain; otherwise, they are included in the alphabetics domain (phonics construct). This domain does not have constructs.

Domain name	Construct name and description
<i>Writing productivity</i>	Includes measures that focus solely on writing quantity such as counts of written words (including different words or content-specific words), sentences, ideas, or length of passages. For example, measures of writing fluency that count the number of words produced in a particular length of time or the number of content-specific words in students' writing samples fall within this domain. This domain does not have constructs.
<i>Writing quality</i>	<p>Writing effective, clear and well-organized text. Assessments of writing quality could focus on writing generally or on particular kinds of writing (such as narrative, exposition, or argument). They can also evaluate a single or multiple components of writing skills (such as complexity or variation in words or sentence structure used, quality or richness of ideas, use of appropriate genre elements, organization of ideas, elaboration of ideas, effectiveness of a story or argument, style or voice, and overall writing quality). For example, state writing assessments in which students write one or more essays under standardized conditions, and automated writing assessments (AWAs) that provide an estimate of overall writing quality fall within this domain.</p> <p>Writing quality outcomes could also include measures of writing conventions and productivity if they include an evaluation of text written by students and are combined with other aspects of writing quality discussed above. This domain does not have constructs.</p>
<i>General literacy achievement</i>	Outcomes that combine separate measures of two or more of the previous domains by providing a summary score across constructs, such as a "total reading score" on a standardized reading test or an "English Language Arts" score on a state test that includes both reading and writing subscores. This domain does not have constructs.

Domain name	Construct name and description
<i>English language proficiency</i>	<p>The English language proficiency domain includes outcomes in the areas of vocabulary (including academic language), oral language, listening comprehension, and grammar.</p> <p>Vocabulary—measures understanding the meanings of words (receptive vocabulary) and using words appropriately (expressive vocabulary). It includes the understanding of <i>academic language</i>, which is the language used for formal discourse in academic disciplines such as mathematics, literature, economics, science, and history. Here we consider terms that cross disciplines (such as “in contrast,” “permutation,” “enable,” “facilitate,” and “comprehensive”) to be part of academic language, but terms that are unique to one discipline (such as “hypotenuse,” “thermodynamics,” “angular momentum”) to be part of the relevant content domain.</p> <p>Oral language—refers to listening and speaking skills.</p> <p>Listening comprehension—refers to understanding spoken language.</p> <p>Grammar—refers to the appropriate use of language (spoken or written) in terms of syntax (sentence structure) or morphology (word inflections).</p>
<i>Mathematics achievement</i>	<p>Mathematics achievement includes outcomes in the following areas: math facts, number sense, number and operations, fractions, measurement, data analysis, algebra, geometry, word problems and general mathematics achievement.</p>
<i>Science achievement</i>	<p>Science achievement includes outcomes in the following areas: science facts, and the capacity to use the tools, procedures, inquiry, nature of science, argumentation in science, and reasoning processes of science. This includes subjects such as biology, chemistry, and earth science.</p>
<i>Social studies achievement</i>	<p>Social studies achievement includes, but is not limited to, outcomes in areas included in the National Assessment of Educational Progress (NAEP), which measures knowledge of US history, world history, geography, economics, government, and civics.</p>

Notes: ^aThe term *phonics* also refers to an instructional approach that focuses on the correspondence between sounds and symbols and is often used in contrast to whole language instructional approaches. For the purposes of defining eligible outcome measures, we use the term *phonics* as defined above, not as an instructional approach.

February 2020 Revisions to Outcome Domains

The WWC released an English Learners review protocol to be used in conjunction with the *WWC Procedures and Standards Handbooks (version 4.0)* in February 2019. The outcome domains in this updated version of that protocol have been revised in two ways:

1. This updated protocol changes the name of the comprehension domain that appeared in the original release of the protocol to be the reading comprehension domain to better reflect the narrow content of the outcomes that are eligible within the domain. In both the original release and this revised version, this domain includes reading comprehension outcomes but excludes vocabulary outcomes. In contrast, vocabulary outcomes are included in the comprehension domain in the Adolescent Literacy review protocol (version 4.0) because vocabulary is closely related to other comprehension skills. Because of the importance of vocabulary to English language proficiency, this English Learners review protocol includes vocabulary outcomes in the English language proficiency domain, instead of with reading comprehension outcomes.
2. This updated protocol replaces the general reading domain that appeared in the original release of the protocol with a general literacy domain that includes outcomes that have writing components. Measures that combine reading and writing subscores, such as some state English Language Arts assessments, would have been ineligible for review using the original release of the protocol. This updated protocol allows these measures to be reviewed and more closely aligns the review protocol with the Adolescent Literacy review protocol (version 4.0), which also includes a general literacy domain.

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 Standard 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?”
- 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 based on the assumption that most attrition in studies of English learners was due to factors that were not strongly related to intervention status. For example, these factors may include family mobility or student 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.

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 (such as 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 joiner 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 might 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 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.
 - If the intervention may affect enrollment or placement decisions (such as a highly publicized program for struggling readers), 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 it is unlikely that the intervention affects enrollment or placement decisions (such as a low-profile, school-wide intervention or curriculum [for example, *Success for All*[®] or *Pearson Literature*[®]]), then ***only late joiners pose a risk of bias***. 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 English Learner topic area, 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, the typical scenarios the WWC encounters in cluster RCTs for the English Learners topic area 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, review team leadership has discretion to make a decision on which 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 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 RDDs in Section III.C

1. Baseline equivalence of individuals

Acceptable pre-intervention measures. 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 the pre-intervention characteristics described below and summarized in Exhibit 3. With the exception of reading and writing outcomes in the specific instances explained below, all outcomes in English must use pre-intervention measures in English, and all outcomes in students’ native language must use pre-intervention measures in students’ native language.

Acceptable pre-intervention measures for outcomes in the three reading domains. For outcomes in any of the three reading domains of ***alphabeticity, reading fluency, and reading comprehension***, a pre-intervention measure within the same domain can be used to establish baseline equivalence. If a pre-intervention measure in the same domain is not available, a pre-intervention measure in *any* of the three outcome domains in reading can be used. For example, a pretest from an alphabeticity domain can be used to establish baseline equivalence for a reading fluency outcome

when a reading fluency pretest is unavailable. Reading pretests administered in either English or students' native language can be used to establish equivalence for reading outcomes in either language *if* students' instruction is provided in both languages when the pre-intervention measure is administered *and* the students' native language uses a roman alphabet. In addition, for students in grades K-1 whose native language uses a roman alphabet, a reading pre-intervention measure in their native language is acceptable regardless of the language(s) of instruction. Pre-intervention measures of English language proficiency or writing are not acceptable for establishing equivalence for outcomes in any of the reading domains since they have not been shown to have strong associations with reading.

Acceptable pre-intervention measures for outcomes in the three writing domains. For outcomes in the three writing domains of writing conventions, writing productivity, and writing quality, a pre-intervention measure from the *same domain* as the outcome can be used to establish baseline equivalence. Writing pretests administered in either English or students' native language can be used to establish equivalence for writing outcomes in either language *if* students' instruction is provided in both languages when the pre-intervention measure is administered *and* the students' native language uses a roman alphabet.

Acceptable pre-intervention measures for outcomes in the general literacy achievement domain. For general literacy achievement outcomes, baseline equivalence rules follow the rules above depending on whether the outcome measure includes only reading or writing skills. (That is, if the outcome measure included subscores in reading domains, baseline equivalence will follow the rules above for the reading domains and if the outcome measure included subscores in writing domains, baseline equivalence will follow the rules above for the writing domains.) If a general literacy outcome measure combines reading and writing skills, then equivalence must be established on *both* a pre-intervention reading measure (from any reading domain) and pre-intervention writing measures in the same domains as the writing skills measured by the outcome.

Acceptable pre-intervention measures for outcomes in the English language proficiency domain. For outcomes in the English language proficiency domain, a pre-intervention measure from the *same domain* as the outcome can be used to establish baseline equivalence.

Acceptable pre-intervention measures for outcomes in the mathematics achievement domain. For outcomes in the mathematics achievement domain, a pre-intervention measure from the *same domain* as the outcome can be used to establish baseline equivalence.

Acceptable pre-intervention measures for outcomes in the science achievement domain. For outcomes in the science achievement domain, a pre-intervention measure in the science achievement, mathematics, reading comprehension, or general literacy achievement domains can be used to establish baseline equivalence.

Acceptable pre-intervention measures for outcomes in the social studies achievement domain. For outcomes in the social studies achievement domain, any pre-intervention measure in the social studies achievement, reading comprehension, or general literacy achievement domains can be used to establish baseline equivalence.

Exhibit 3. Acceptable Pre-Intervention Measures by Outcome Domain

Outcome domain	Acceptable pre-intervention measures
<ul style="list-style-type: none"> • Alphabetics • Reading fluency • Reading comprehension 	<ul style="list-style-type: none"> • A pre-intervention measure in <i>any</i> of the three reading domains
<ul style="list-style-type: none"> • Writing conventions • Writing productivity • Writing quality • English language proficiency • Mathematics achievement 	<ul style="list-style-type: none"> • A pre-intervention measure in the <i>same</i> domain as the outcome
<ul style="list-style-type: none"> • General literacy achievement 	<ul style="list-style-type: none"> • If the outcome measure includes only reading skills, a pre-intervention measure in any of the three reading domains • If the outcome measure includes only writing skills, a pre-intervention measure in each of the writing domains represented in the measure • If the outcome includes both reading and writing skills, a pre-intervention measure in any reading domain and a pre-intervention measure in each of the writing domains represented in the measure
<ul style="list-style-type: none"> • Science achievement 	<ul style="list-style-type: none"> • A pre-intervention measure in the science achievement, mathematics achievement, reading comprehension, or general literacy achievement domain
<ul style="list-style-type: none"> • Social studies achievement 	<ul style="list-style-type: none"> • A pre-intervention measure in the social studies achievement, reading comprehension, or general literacy achievement domain

Baseline effect sizes greater than 0.25 standard deviations

For this review, the WWC will measure baseline differences between the intervention and comparison groups for each pre-intervention measure reported in the study that are listed above in Exhibit 3 as acceptable for establishing baseline equivalence for the outcome measure. In general, when the baseline effect size for the analytic sample is larger than 0.25 standard deviations for any acceptable pre-intervention measure, findings for all outcome measures in the outcome domain using the same analytic sample are rated *Does Not Meet WWC Group Design Standards*.

For example, for the three reading domains of ***alphabetics, reading fluency, and reading comprehension,*** this review will measure the baseline difference between intervention and comparison groups for each pre-intervention measure reported in the study within the eligible reading domains. If there is a baseline effect size greater than 0.25 standard deviations in the

analytic sample for *any* pre-intervention measure in one of the reading domains, analyses of outcomes in *all* three reading domains using the analytic sample are rated *Does Not Meet WWC Group Design Standards*.

Similarly, for the *science achievement* domain, this review will measure baseline differences between intervention and comparison groups for the eligible pre-intervention measures that are specified above as acceptable measures for establishing baseline equivalence, i.e., science achievement, mathematics achievement, reading comprehension, or general literacy achievement. If there is a baseline effect size greater than 0.25 standard deviations in the analytic sample for *any* of these pre-intervention measures, analyses of all outcomes *within* the science achievement domain using the analytic sample are rated *Does Not Meet WWC Group Design Standards*.

Required statistical adjustments

For the three reading domains of *alphabeticity, reading fluency, and reading comprehension*, when the baseline difference for the analytic sample is in the statistical adjustment range (between 0.05 and 0.25 standard deviations) for any pre-intervention measure *within* a domain, all outcome measures 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 analytic sample, 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*. When baseline equivalence for an outcome in any of these domains is assessed based only on pre-intervention measures from outside the domain of the outcome measure, the team leadership, in consultation with content expert, has discretion to decide which measures must be included as statistical adjustments.

For the remaining domains in this protocol, this review will determine if the baseline difference for the analytic sample is in the statistical adjustment range for *any* pre-intervention measure listed above in Exhibit 3 as acceptable for establishing baseline equivalence for the outcome measure (either from within the same domain as the outcome or from another domain). If so, analyses of all outcomes within the same domain using the same analytic sample must adjust for those baseline differences. For example, in an analysis where the outcome is in the social studies domain, if the pre-intervention differences for both a general literacy achievement measure and a social studies measure are in the statistical adjustment range, the analysis of all social studies achievement outcome measures must include statistical adjustments for the pre-intervention general literacy achievement and social studies measures.

Additional baseline equivalence characteristics

In addition to the pre-intervention measures that are required for satisfying the baseline equivalence requirement, baseline equivalence should be established for other sample characteristics including (1) English language proficiency, (2) student age or grade level, (3) measures of students' socio-economic status (such as free or reduced-price lunch status, parental education, family income, or poverty status), (4) English learner (EL) status, and (5) special education or disability status, if they are reported by the study. 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 are larger than 0.25 standard deviations for continuous measures (such as age) or dichotomous measures (such as EL status), 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 on characteristics in question.

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 listed above for assessing baseline equivalence of individuals, and the same statistical adjustment requirements apply.

Acceptable samples for demonstrating baseline equivalence of clusters

For the English learners topic area, any of the following three sources 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). In addition, the sample used for baseline equivalence (either 2 or 3 below) also has to be the EL subgroup or include at least 50 percent ELs for findings on the full sample.

- 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, a reviewer should assess baseline equivalence using the sample listed first in the list above—that is, (1) should be used if available, then (2), and then (3). 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 are more stringent than those specified in the *WWC Standards Handbook* (in Section IV.A: Outcome Requirements and Reporting). Specifically, this review requires a minimum of 0.60 for internal consistency (as measured by, for example, Cronbach's alpha) of the alphabetics, reading fluency, reading comprehension, general literacy achievement, writing conventions, writing productivity, writing quality, English language proficiency, and social studies outcome measures; and 0.50 for internal consistency of mathematics and science achievement outcome measures. The default threshold for acceptable internal consistency in the *WWC Procedures Handbook (version 4.0)* is 0.50. The higher internal consistency requirements for literacy measures are aligned with the Adolescent Literacy topic area protocol.

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 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 will refine the potential scope of this search by including additional search terms, such as English learner, limited English proficient, English language learner, non-English speaker, English as a Second Language, ESL, English for Speakers of Other Languages, ESOL, language minority, second language learner, and dual language learner.

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:

- Center for Applied Linguistics
- Center for Comprehensive School Reform and Improvement
- Center for Data-Driven Reform in Education (CDDRE) at Johns Hopkins University
- Center for Research in Educational Policy (CREP)
- Center for Research on Education, Diversity and Excellence (CREDE)
- Center on Instruction
- Harvard Graduate School of Education
- Johns Hopkins University School of Education
- The Meadows Center for Preventing Educational Risk (University of Texas at Austin)
- Mid-Continent Research for Education and Learning
- National Science Foundation (NSF)
- National Science Teachers' Association (NSTA)
- Public Education Network
- Southwest Educational Development Laboratory (SEDL)
- The University of California, Los Angeles (UCLA)
- Understanding Language (Stanford University)
- Wisconsin Center for Education Research (WCER)
- WestEd

REFERENCES

- National Center for Education Statistics (2014). Table 1. Public high school 4-year adjusted cohort graduation rate (ACGR), by race/ethnicity and selected demographics for the United States, the 50 States, and the District of Columbia: School year 2013–14. Retrieved from: https://nces.ed.gov/ccd/tables/ACGR_RE_and_characteristics_2013-14.asp.
- U.S. Department of Education, Office of English Language Acquisition, Language Enhancement, and Academic Achievement for Limited English Proficient Students (2018). The biennial report to Congress on the implementation of the Title III State Formula Grant Program, school years 2012–14, Washington, D.C.