

WWC EVIDENCE REVIEW PROTOCOL FOR INTERVENTIONS FOR ENGLISH LANGUAGE LEARNERS VERSION 2.1

Topic Area Focus

The What Works Clearinghouse (WWC) review focuses on interventions designed to improve the English language literacy and/or academic achievement of elementary and middle school students (grades K–8) who are English language learners (ELLs).

Systematic reviews of evidence in this topic area will address the following question:

- Which programs for elementary and middle school English language learners increase the English language or academic outcomes (reading or math) of these students?

Key Definitions

English Language Learners (ELLs). ELLs are students with a primary language other than English who have a limited range of speaking, reading, writing, and listening skills in English. This might include students who have been identified and determined by their school as having limited English proficiency (LEP)

- at the time of the study, or
- within the preceding two years.

Terms such as Limited English Proficiency (LEP), English Learners (EL), non-English Speakers, English as a Second Language (ESL), English for Speakers of Other Languages (ESOL), Language Minority (LM), or Second Language Learners (SLL) may also appear, and should be brought to the attention of the principal investigator (PI) for determination of eligibility.

English Language Skills. These skills include speaking, listening, reading, and writing in English.

Academic Domains. Student achievement in the domains of *reading*, *mathematics*, and *English language development* are of interest.

Reading outcomes include measures of:

- word reading,
- fluency and/or accuracy in reading connected text,
- vocabulary,
- reading comprehension, and
- general reading achievement.

Specifically, reading outcomes can include pseudo-word reading tasks, but cannot include early reading-related skills, such as rhyming, phonemic awareness, and letter naming. Spelling outcomes are not considered reading outcomes.

Mathematics outcomes include measures of:

- concepts and procedures,
- word problems and applications, and
- general math achievement.

English language development outcomes include measures of:

- listening comprehension,
- receptive vocabulary,
- grammar, syntax, and
- other linguistic features of English.

ELIGIBILITY CRITERIA AND EVIDENCE STANDARDS

Populations to be Included

This WWC review includes elementary and middle school students (grades K–8) who are English language learners.

Types of Interventions That May be Included

Only research on interventions that are replicable (i.e., documented well enough that they can be reproduced) will be reviewed.

The interventions eligible for review could include the following:

Programs. Programs considered for this study will *typically* consist of educational content that extends over the course of a semester or more of instruction. Programs may be based on text materials, computer software, videotapes, professional development packages for teachers, or any other material base. For purposes of this review, a program is thought of as a replicable, materials-based instructional program for which:

- the population of learners is well-specified (e.g., linguistic background, academic skills),
- learning goals are well-delineated, and
- assessments provide clear information about progress toward those learning goals.

Instructional Practices. Interventions that focus on instructional strategies (i.e., educational practices, such as pre-teaching vocabulary words or cooperative learning strategies) will be included in this review.

The following types of studies are excluded:

- studies that compare differing languages of instruction (e.g., teaching first graders in Spanish vs. English), or
- studies where all instruction is conducted in the students' native language.

Much of the academic ELL literature focuses on the point when it is best for ELLs to be fully immersed in English instruction in lieu of teaching them in their native language (August & Hakuta, 1997; Gersten & Baker, 2000). The report of the National Research Council (August & Hakuta, 1997) concluded that these evaluations are problematic and do not provide an empirical basis for determining federal or local policy. We concur with this appraisal. Therefore, studies that use a different language (or mix of languages) in intervention and comparison conditions are ineligible for the review; some allowance will be made for investigations that had minor degrees of differing language use across conditions if this

resulted from implementation problems only (i.e., not purposeful manipulation).¹ For example, a study that compares the effects of transitional bilingual educational (TBE) approaches versus structured immersion falls outside the scope of this review. However, studies that compare different forms of TBE are of interest.

A study of teaching reading or mathematics in a students' native language is a legitimate study, but does not provide information on how to deal with the challenging task of teaching academic material to ELL students using a language they have not yet mastered.

The review team will include studies where the majority of instruction is in English but where up to 20% of instruction is in the students' native language. Many educators advocate that when ELLs receive academic instruction in English, some level of strategic native language support is advisable. We therefore included interventions where the majority of the instruction is conducted in English but teachers occasionally provide some native language support. We chose 20% as a reasonable estimate of the differences between occasional native language and an approach with a substantial native language or bilingual component.

Interventions conducted in afterschool programs are not included in this review.

Types of Studies That Will be Included

Design. The review focuses on empirical studies of intervention effectiveness using quantitative methods and inferential statistical analyses. These designs include randomized controlled trials, a regression discontinuity design, a quasi-experimental design, or a single-case design.²

Publication Date. This review is limited to empirical studies published in 1983 or later.

Sample. The studies must include students in grades K through 8 in the United States, its territories, or tribal entities. The study sample must have a subgroup analysis for ELLs, or the sample must include at least 60% ELLs. The PI will consult with the PI of the Learning Disability topic area to determine when studies that include students with learning disabilities should be reviewed by the ELL topic area or the Learning Disability topic area.

Outcome Measures. The outcome measures must measure student achievement, not teacher outcomes, and at least one of the outcome measures needs to be academic in nature. The measures must demonstrate sufficient reliability or face validity according to WWC standards. Outcome measures must also be administered immediately after the intervention, or no more than a few months after the intervention was concluded.

¹ Implementation problems will be noted in WWC reports.

² The WWC is in the process of developing standards for reviewing and reporting on regression discontinuity or single-case design studies. Consequently, studies using these designs will be included in the review when the standards become available. They are included in the report references with a note indicating that standards are not yet available for that research design.

Specific Topic Parameters

Commonly-shared or theoretically-derived characteristics of the intervention

For studies that meet standards, reviewers describe the information provided regarding intervention implementation. In the report, the review may consider:

- Student native language proficiency
- Student English language proficiency at pretest
- Other academic pretest measures
- Length of intervention
- Intensity of intervention (i.e., number of hours of instruction)
- Description of the intervention (program or practice)
- Teacher training in intervention strategy
- Degree of program implementation/fidelity
- Context of instruction (e.g., special education, ESL class, regular classroom, dual immersion classroom etc.)
- Material such as texts, videotapes, software or other classroom materials
- ELL definition used by the school or researcher
- Eligible outcome measures

If available, other important characteristics include:

- Required training needed to carry out the intervention
- Use of support materials and prescribed classroom structures
- Reference for widely available curricula (e.g., a commercially available supplement for ELLs from a core reading program, a commercially available ESL or English Language Development curriculum)
- Descriptions of guiding principles informing the interventions, practices, and programs will also be considered

Important characteristics of the intervention for replication

- An intervention is replicable if the intervention is “branded”; that is, commercial programs and products. The intervention must possess any of the following characteristics:
 - An external developer provides technical assistance, sells or distributes the intervention.
 - The intervention is packaged or otherwise available for distribution/use beyond a single site.
 - The intervention is trademarked.
- An intervention is replicable if the intervention is “not branded” but meets the following conditions:

- The intervention is described in general terms.
- The duration of the intervention is described.
- The characteristics of the intervention are described.
- The target population is identified.
- The curriculum and/or instructional practices used are described.

Eligible outcomes classes

Achievement in the domains of *reading*, *mathematics*, and *English language development* are of interest.

Reading outcomes include measures of:

- word reading,
- fluency and/or accuracy in reading connected text,
- vocabulary,
- reading comprehension, and
- general reading achievement.

Specifically, reading outcomes can include pseudo-word reading tasks, but cannot include early reading-related skills such as rhyming, phonemic awareness, and letter naming. Spelling outcomes are not considered reading outcomes.

Mathematics outcomes include measures of:

- concepts and procedures,
- word problems and applications, and
- general math achievement (i.e. a standardized test covering a full array of mathematics topics).

English language development includes measures of oral or written proficiency including measures of:

- listening comprehension,
- receptive vocabulary,
- grammar, syntax and
- other linguistic features of the English language.

Nationally normed tests, standardized tests, and researcher developed measures are considered eligible in any of these three domains.

Reliability of Eligible Outcome Measures

WWC standards for reliability considerations are as follows:

- Internal consistency: minimum of 0.60

- Temporal stability/test-retest: minimum of 0.40
- Inter-rater reliability: minimum of 0.50

Interval of time that studies should have been conducted

Studies must have been conducted or published since 1983 (i.e., with a publication date of 1983 or later). This is the default time interval for all WWC reviews. The decision to maintain the 1983 cutoff was made after consulting with a number of experts in the ELL field.

Necessary characteristics of the target population

ELL Status. ELLs are students with a primary language other than English who have a limited range of speaking, reading, writing, and listening skills in English. This might include students who have been identified and determined by their school as having limited English proficiency

- at the time of the study, or
- within the preceding two years.

Terms such as Limited English Proficiency (LEP), English Learners (EL), non-English Speakers, English as a Second Language (ESL), English for Speakers of Other Language (ESOL), Language Minority (LM), or Second Language Learners (SLL) may also appear, and should be brought to the attention of the PI for his determination of eligibility.

Grade. ELL students must be in grades K through 8.

Location. Students reside and attend a school within the United States (including U.S. territories and tribal entities).

Equivalence

If the study design is a randomized controlled trial (RCT) 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. Sufficient reporting of pre-intervention data should be included in the study report (or obtained from the study authors) to allow the review team to draw conclusions about the equivalence of the intervention and comparison groups. Pre-intervention characteristics can include the outcome measure(s) administered prior to the intervention or other measures that are not the same as, but are highly related to, the outcome measure(s). For the ELL review, the variables on which studies must demonstrate equivalence are: pretest scores or a reasonable predictor of posttest performance, grade level, and the level of English language skills. The next section describes these characteristics in more detail.

Groups are considered equivalent if the reported differences in pre-intervention characteristics of the groups are less than or equal to one-quarter of the pooled standard deviation in the sample, regardless of statistical significance. However, if differences are greater than 0.05 standard

deviations and less than or equal to one-quarter of the pooled standard deviation in the sample, the analysis must control analytically for the individual-level pre-intervention characteristic(s) on which the groups differ. If there are pre-intervention differences greater than 0.25 for *any* of the listed characteristics, the study does not meet standards. In addition, if there is evidence that the populations were drawn from very different settings (such as rural versus urban, or high socioeconomic status [SES] versus low-SES), the PI may decide that the environments are too dissimilar to provide an adequate comparison.

Important characteristics for equating groups of participants in a study

Important characteristics of participants that might be related to the intervention's effect and must be equated include:

- Pretest scores for at least one outcome measure or a reasonable predictor of posttest performance.
 - Measures of Spanish language should be considered to establish equivalence if provided and the use of the assessment logically fits within the purpose and context of the study.
 - Measures like phonemic awareness and alphabetic knowledge in English and Spanish can also be used to establish equivalence for grade K through 1 studies.
 - For the English language domain, oral and/or written language must be documented as equivalent. Language measures can be used to demonstrate equivalence for oral language development studies.
 - For the mathematics domain, measures of mathematics in English and Spanish can be used for equivalence.
- Grade level
- Level of English language skills

Interval for outcome measures

The benefits of ELL interventions are expected to appear by the end of the intervention. Thus, measures at the end of an intervention are appropriate. Delayed outcomes are not considered in this protocol, as we want to ensure that readers can make appropriate comparisons among programs, and so few programs study delayed effects.

Attrition in RCTs

The WWC considers both the overall sample attrition rate and the differential in sample attrition between the intervention and comparison groups, as both contribute to the potential bias of the estimated effect of an intervention. The WWC has established conservative and liberal standards for acceptable levels of attrition. The conservative standards are applied in cases where the PI has reason to believe that much of the attrition is endogenous to the intervention reviewed—for example, high school students choosing whether or not to participate in an after school program. The liberal standards are applied in cases where the PI has reason to believe that much of the attrition is exogenous to the intervention reviewed

(e.g., in cases where movement of young children in and out of school districts is due to family mobility). Attrition rates are based on the number of sample cases used in the analysis sample with measured (as opposed to imputed) values of the outcome measures.

The English Language Learners review uses the liberal standard, reflecting the assumption that most attrition in studies of ELLs is due to factors that are not strongly related to intervention status, such as parent mobility and absences on the days that assessments are conducted. Table 1 presents the maximum difference in the attrition rate for the intervention and comparison group that is acceptable for a given level of overall sample attrition. The empirical basis for these thresholds is described in Appendix A of the *WWC Procedures and Standards Handbook*, version 2.1.

Studies based on cluster random assignment designs must meet attrition standards for both the study sample units that were assigned to intervention or comparison group status (e.g., schools or districts) and the study sample units for analysis (e.g., typically, students). In applying the attrition standards to the subcluster level (e.g., students), the denominator for the attrition calculation includes only sample members in the clusters that remained in the study sample. RCTs with combinations of overall and differential attrition rates that exceed the applicable threshold, based on the applicable standard, must demonstrate baseline equivalence of the analysis sample or, if nonequivalence falls within the allowable range, statistically control for the nonequivalence, in order to receive the second-highest rating: *meets WWC evidence standards with reservations*. See the Baseline Equivalence section for more details.

**Table 1: Attrition Standards for Randomized Controlled Trials
Highest Level of Differential Attrition Allowable to Meet the Attrition Standard
Under the Liberal Attrition Standard**

Overall Attrition	Allowable Differential Attrition	Overall Attrition	Allowable Differential Attrition
0	10.0	34	7.4
1	10.1	35	7.2
2	10.2	36	7.0
3	10.3	37	6.7
4	10.4	38	6.5
5	10.5	39	6.3
6	10.7	40	6.0
7	10.8	41	5.8
8	10.9	42	5.6
9	10.9	43	5.3
10	10.9	44	5.1
11	10.9	45	4.9
12	10.9	46	4.6
13	10.8	47	4.4
14	10.8	48	4.2
15	10.7	49	3.9
16	10.6	50	3.7
17	10.5	51	3.5
18	10.3	52	3.2
19	10.2	53	3.0
20	10.0	54	2.8
21	9.9	55	2.6
22	9.7	56	2.3
23	9.5	57	2.1
24	9.4	58	1.9
25	9.2	59	1.6
26	9.0	60	1.4
27	8.8	61	1.1
28	8.6	62	0.9
29	8.4	63	0.7
30	8.2	64	0.5
31	8.0	65	0.3
32	7.8	66	0.0
33	7.6	67	-

Statistical and analytical issues

RCT studies with low attrition do not need to use statistical controls in the analysis, although statistical adjustment for well-implemented RCTs is permissible and can help generate more precise effect size estimates. For RCTs, the effect size estimates will be adjusted for differences in pre-intervention characteristics at baseline (if available) using a difference-in-differences method if the authors did not adjust for pretest (see Appendix B of the *Handbook*). Beyond the pre-intervention characteristics required by the equivalence standard, statistical adjustment can be made for other measures in the analysis as well, though they are not required.

For the WWC review, the preference is to report on and calculate effect sizes for post-intervention means adjusted for the pre-intervention measure. If a study reports both unadjusted and adjusted post-intervention means, the WWC review will report the adjusted means and unadjusted standard deviations. If adjusted post-intervention means are not reported, they will be requested from the author(s).

The statistical significance of group differences will be recalculated if (a) the study authors did not calculate statistical significance, (b) the study authors did not account for clustering when there is a mismatch between the unit of assignment and unit of analysis, or (c) the study authors did not account for multiple comparisons when appropriate. Otherwise, the review team will accept the calculations provided in the study.

When a misaligned analysis is reported (that is, the unit of analysis in the study is not the same as the unit of assignment), the effect sizes computed by the WWC will incorporate a statistical adjustment for clustering. The default intraclass correlation used for the ELL review is 0.20 for cognitive, language, literacy, and math outcomes. For an explanation about the clustering correction, see Appendix C of the *Handbook*.

When multiple comparisons are made (that is, multiple outcome measures are assessed within an outcome domain in one study) and not accounted for by the authors, the WWC accounts for this multiplicity by adjusting the reported statistical significance of the effect using the Benjamini-Hochberg correction. See Appendix D of the *Handbook* for the formulas the WWC uses to adjust for multiple comparisons.

All standards apply to overall findings as well as analyses of sub-samples.

LITERATURE SEARCH METHODOLOGY

Literature Search Strategies

The WWC literature search is comprehensive and systematic. Detailed protocols guide the entire literature search process. At the beginning of the process, relevant journals, organizations, and experts are identified. The WWC searches core sources and additional topic-specific sources identified by the PI and the Senior Content Advisor. The process is fully and publicly documented. Trained WWC staff members locate and collect studies.

2005 Literature Search

This section contains topic specific elements of the literature search (e.g., search terms, additional journals, and associations) performed in between 2003 and 2005. The final section describes an expanded search conducted in 2009. In 2005, the ELL team searched for studies evaluating the effectiveness of ELL interventions published from 1983 through 2005. In 2009, the team searched for studies published since 2003.

Keyword List

Bilingual Cooperative Integrated Reading and Composition (BCIRC)	Limited English proficient students
Bilingual education	Pull out ESL programs
Cognitive Academic Language Learning Approach (CALLA)	Second language acquisition
Dual immersion	Second language education
Dual language	Sheltered English
English as a second language	Sheltered immersion
English language learners	Sheltered instruction
English learners	SRA Reading Mastery
ESL students	Structured immersion
High intensive language training	Success for All (Studies with outcomes pertaining to ELL students)
Immersion programs	Targeted English
Into English	

Journals

The Cochrane Collaboration and the Campbell Collaboration have regarded hand searching of journals as the gold standard in retrieving studies. The yields obtained from hand searches are usually more than from electronic database searches. For a comprehensive review of the literature, each and every article in the journal is examined, even though this is a tedious and time-consuming process. Below we list the topic-specific journals used for the English language learners hand searches:

American Educational Research
Journal
Bilingual Research Journal
Elementary School Journal
Evaluation and Research in
Education
Exceptional Children
Journal of Educational Issues for
Language Minority Students
Journal of Educational Psychology

Journal of Learning Disabilities
Journal of Multilingual and
Multicultural Development
Journal of Special Education
Language, Culture, and Curriculum
Learning Disabilities Research and
Practice
Remedial and Special Education
TESOL Quarterly

Supplementary List of English Language Learners Organizations

Linguistic Society of America (LSA)
National Association for Bilingual Education (NABE)
Teachers of English to Speakers of Other Languages (TESOL)

Personal Contacts

The ELL team solicited studies directly from experts, identified by the PI, in the field of education who work on ELL interventions. Another source of contacts was individuals identified using listservs dedicated to ELL, whose members are scholars working in this area.

2009 Literature Search³

Keyword List

The 2009 literature search contained the keywords searched in 2005, as well as the keywords listed below.

Arthur TV program	NEARStar
Augmenting Thinking through Language	On Our Way to English
Acquisition skills	Peer tutoring and response groups
Curriculum-based instruction	Pre-teaching vocabulary
Effective use of time	Proactive reading
Enhanced proactive reading	Project MASTER
ESL in the content areas	Read Naturally
Front row phonics	Read Well
Hampton-Brown	Reading Recovery
Instructional conversations and literature	Second language
Logs	Sheltered Instruction Observation Protocol

³ In 2005, the WWC searched for ELL studies including students in grades K–6. Because the team has expanded its review to include seventh- and eighth-grade students, the MPR library conducted a retroactive search using the 2009 literature search keywords, databases, and targeted research websites to assure that all studies containing students in grades K–8 published between 1983 and 2009 are included in the review.

Instructional practices
Lectura Proactiva
Metacognitive teaching approaches

Vocabulary improvement
Vocabulary improvement program

In addition to searching the above keywords, we performed specific searches for each of the interventions identified in the 2007 ELL topic report.

A combination of Boolean terms such as AND and OR were used with this keyword list. Libraries at MPR conducted the actual searching and should be consulted as to the appropriate combination to use for searching within each electronic database.

Databases

Academic Search Premier
Business Source Corporate
Campbell Collaboration
Dissertation Abstracts
EconLit
Education Research Complete

EJS E-Journals
ERIC
Google Scholar
PsycINFO
SocINDEX with Full Text
WorldCat

Targeted Research Websites

Abt Associates
Alliance for Excellent Education
American Enterprise Institute
American Institutes of Research
American Speech-Language-Hearing Association (ASHA)
Appalachian Education Laboratory (Edvantia)
Best Evidence Encyclopedia
Broad Foundation (Education)
Brookings Institution
Carnegie Corporation of New York
Center for Comprehensive School Reform and Improvement
Center for Data-Driven Reform in Education
Center for Research and Reform in Education
Center for Research in Educational Policy (CREP)
Center for Social Organization of Schools
Center on Education Policy
Center on Instruction
Chapin Hall Center for Children at the University of Chicago

MDRC
Mid-continent Research for Education and Learning
National Association for Bilingual Education (NABE)
National Association of State Boards of Education
National Autism Center—National Standards Project
National Center on Secondary Education and Transition
National College Access Network
National Dissemination Center for Children with Disabilities
National Dropout Prevention Center/Network
National Governors' Association
National Reading Panel
Pacific Resources for Education and Learning (PREL)
Pathways to College Network
Public Education Network
Public Policy Research Institute at Texas A&M University
Public/Private Ventures (PPV)

Congressional Research Service (via
OpenCRS.org)
Florida Center for Reading Research
(FCRR)
Government Accountability Office (GAO)
Harvard Graduate School of Education
Heritage Foundation
Hoover Institution
Institute for Higher Education Policy
Institute for Public Policy and Social
Research (IPPSR)
Johns Hopkins University School of
Education
Learning Point Associates
Linguistic Society of America (LSA)
Mathematica Policy Research

RAND
Southwest Educational Development
Laboratory (SEDL)
SRI
Teachers of English to Speakers of Other
Languages (TESOL)
Technical Assistance Center on Social
Emotional Intervention for Young
Children
The Education Resources Institute
Thomas B. Fordham Institute
U.S. Department of Education (includes
Institute for Education Sciences,
National Center for Special Education
Research etc)
Urban Institute

Additionally, if the developer of an intervention listed in the 2007 ELL topic report has a website, the ELL team searched the site for studies evaluating the effectiveness of the intervention published since 2003.

References

- August, D., & Hakuta, K. (Eds.). (1997). *Improving schooling for language-minority children: A research agenda*. Washington, DC: National Academy Press.
- Gersten, R., & Baker, S. (2000). What we know about effective instructional practices for English-language learners. *Exceptional Children*, 66(4), 454–470.