EVIDENCE REVIEW PROTOCOL FOR INTERVENTIONS FOR ENGLISH LANGUAGE LEARNERS, VERSION 2.2

Topic Area Focus

The What Works Clearinghouse (WWC) review in the English language learners (ELLs) topic area 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 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, mathematics, science, or social studies) 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 review team leadership for determination of eligibility.

English language skills. These skills include speaking, listening, reading, and writing in 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. Students must reside and attend a school within the United States (including US territories and tribal entities). The study sample must have a subgroup analysis for ELLs, or the sample must include at least 50% ELLs. The review team leadership of the ELL topic area will consult with the review team leadership of the Students with Learning Disabilities topic area to determine whether studies that include students with learning disabilities should be reviewed by the ELL topic area or the Students with Learning Disabilities topic area.

Types of Interventions to be Included

Only research on interventions that are replicable (i.e., documented well enough that they can be reproduced) will be reviewed. The types of interventions eligible for review include the following:

Programs. Curricula or instructional 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 defined as a replicable, materials-based instructional program.

Instructional practices or strategies. An instructional practice or strategy is 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.

Language of instruction. Studies where all instruction for the respective intervention is conducted in the students' native language will be excluded from this topic area review. The rationale is that a study of teaching reading or mathematics in a student's native language does not provide information on how to deal with the challenging task facing schools of teaching academic material to ELL students using a language they have not yet mastered.

Studies of interventions in which the majority of instruction is in English, but where up to 20% of instruction is in the students' native language are eligible, however. Many educators advocate that when ELLs receive academic instruction in English, some level of strategic native language support is advisable. Therefore, interventions where the majority of the instruction is conducted in English, but teachers occasionally provide some native language support, are eligible for review under this protocol. The review team leadership chose 20% as a reasonable estimate of the differences between occasional native language and an approach with a substantial native language or bilingual component.

Studies that examine the effectiveness of instruction in one's native language vs. immersion in English are also not eligible for review under this protocol. For example, a study that compares the effects of transitional bilingual educational (TBE) approaches vs. structured immersion fall outside the scope of this review. The report of the National Research Council (August & Hakuta,

1997) concluded that evaluations of this nature are problematic and do not provide an empirical basis for determining federal or local policy. 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]

Types of Studies to be Included

Design. The review focuses on empirical studies of intervention effectiveness using quantitative methods and inferential statistical analyses. Randomized controlled trials (RCTs) and quasi-experimental designs (QEDs) will be reviewed under the WWC evidence standards for those designs. Regression discontinuity design (RDD) and single-case design (SCD) studies will be reviewed under the pilot standards for those designs.

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

Eligible Outcomes

Studies must have a measure of student academic achievement in one or more of the outcome domains specified below to be eligible for review. The measures must demonstrate sufficient reliability or face validity according to WWC guidelines.

The outcome domains for this review protocol are: reading, mathematics, science, social studies, and English language development.

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 mathematics achievement.

---

[1] Implementation problems will be noted in WWC reports.

Science outcomes include measures of:

- content,

- practices, and

- skills.

Social studies outcomes measure a variety of content knowledge and concept application in the following areas:

- history,

- geography,

- economics, and

- government and civics.

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

- academic language—this term refers to the language used for formal discourse in academic disciplines such as mathematics, literature, economics, science, and history. Here we consider terms that traverse disciplines (e.g., "in contrast," "permutation," "enable," "facilitate," and "comprehensive") to be part of academic language, but terms that are unique to one discipline (e.g., "hypotenuse," "thermodynamics," "angular momentum") to be part of the relevant content domain.

- 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 domains.

Reliability of Eligible Outcome Measures

WWC guidelines for reliability considerations require that a researcher-created measure meets at least one of the following criteria:

- Internal consistency: minimum of 0.60

- Temporal stability/test-retest: minimum of 0.40

- Inter-rater reliability: minimum of 0.50

Interval for outcome measures. The impact of ELL interventions on students' outcomes are expected to appear by the end of the intervention. Thus, outcomes that are measured at the end of an intervention will be used to determine the intervention rating. Long-term impacts will be

presented in a supplemental appendix if those analyses meet WWC standards with or without reservations.

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 review team leadership 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 review team leadership 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 in grades K–8 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 groups that is acceptable for a given level of overall sample attrition under the liberal attrition standard. 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.

Baseline equivalence. RCTs with high attrition and all QEDs must demonstrate baseline (that is, pre-intervention) equivalence between the intervention and comparison groups in the analysis sample in order to receive the rating of *meets WWC evidence standards with reservations*. Equivalence is examined on baseline measures of the outcomes or baseline measures that are expected to be highly correlated with these outcomes.

For the ELL review, for the groups to be considered equivalent, the groups must be similar for each outcome on a pretest or a reasonable predictor of posttest performance. Equivalence is established within each outcome domain; however, if youth are dissimilar on grade level, level of English language skills, or **any** measure of English language development, the study *does not meet WWC evidence standards*.

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	-

These are the guidelines for which types of measures can be used to establish equivalence within each domain.

Pretest scores or a reasonable predictor of posttest performance:

- In the reading domain, in addition to reading pretests, measures like phonemic awareness and alphabetic knowledge in English and Spanish can be used to establish equivalence for grade K–1 studies. A pretest difference on any of these measures or on English language development will indicate a lack of equivalence.

- In the mathematics domain, baseline equivalence can only be demonstrated on mathematics pretest. However, a pretest difference on English language development will indicate a lack of equivalence.

- In the science domain, baseline equivalence can be demonstrated on reading or mathematics tests when science tests are not available at baseline. A pretest difference on any of these measures or on English language development will indicate a lack of equivalence.

- In the social studies domain, baseline equivalence can be demonstrated on reading tests when social studies tests are not available at baseline. A pretest difference on any of these measures or on English language development will indicate a lack of equivalence.

- In the English language development domain, English language proficiency measures should be used to demonstrate baseline equivalence.

Groups are considered equivalent if the reported differences in mean baseline characteristics of the groups are less than or equal to 5% of the pooled standard deviation in the sample. If this is the case, the equivalence standard is met, and the study can receive a rating of *meets WWC evidence standards with reservations.* Statistical significance of the difference in means is not considered.

If differences are greater than 5% and less than or equal to 25% of the pooled standard deviation in the sample, the study findings must be based on analytic models that control for the individual-level baseline characteristic(s) on which the groups differ in order to receive a rating of meets WWC evidence standards with reservations. Otherwise, the study is rated does not meet WWC evidence standards.

If baseline differences are greater than 25% of the pooled standard deviation for any of the pretests within a domain, then none of the outcomes within that domain meet the baseline equivalence standard, regardless of whether or not the impacts are estimated using models that control for baseline characteristics.

Finally, when there is evidence that the populations being compared are drawn from very different settings (such as rural vs. urban, or high-SES vs. low-SES), these settings may be deemed too dissimilar by the review team leadership to provide an adequate comparison. In these cases, the study is rated does not meet WWC evidence standards.

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 reading, mathematics, science, social studies, or English language development 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 review team leadership. The process is fully and publicly documented.

Searches Conducted from 2003 to 2005

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 expanded searches conducted in 2009 and 2012. 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)
Bilingual education
Cognitive Academic Language Learning
    Approach (CALLA)
Dual language
Dual immersion
English as a Second Language
English language learners
English learners
ESL students
High intensive language training
Immersion programs

Into English
Limited English proficient students
Pull out ESL programs
Second language acquisition
Second language education
Sheltered English
Sheltered immersion
Sheltered instruction
SRA Reading Mastery
Structured immersion
Success for All (studies with outcomes pertaining to ELL students)
Targeted 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 review team leadership, in the field of education who work on ELL interventions. Another source of contacts was individuals identified using listservs dedicated to ELLs, whose members are scholars working in this area.

2009 Literature Search[2]

Keyword List

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

Acquisition skills
Arthur TV program
Augmenting Thinking through Language
Curriculum-based instruction
Effective use of time
Enhanced proactive reading
ESL in the content areas
Front Row Phonics
Hampton-Brown
Instructional Conversations and Literature Logs
Instructional Practices
Lectura Proactiva
Metacognitive teaching approaches

NEARStar
On Our Way to English
Peer tutoring and response groups
Pre-teaching vocabulary
Proactive Reading
Project MASTER
Read Naturally
Read Well
Reading Recovery
Second language
Sheltered Instruction Observation Protocol
Vocabulary Improvement
Vocabulary Improvement Program

---

[2] 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 Mathematica 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.

In addition to searching the above keywords, we performed specific searches for each of the interventions identified in the 2007 ELL topic report (note: this is no longer available on the WWC website; however, comparable information can be gathered from using the Find What Works tool on the WWC website).

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

2012 Literature Search

The updated 2012 literature search contained keywords searched in 2009, as well as the keywords listed below.

Academic Language Instruction for All
    Students (ALIAS)
Accelerated Reader
ACCESS Math
ACCESS Newcomers
ACCESS Science
Achieve3000 Math
Achieve3000 Science
ALEKS Math
AWARD Reading
Between the Lions
Bilingual Cooperative Integrated Reading
    and Composition
Breakthrough to Literacy
Chechne Konnen Project
Chemistry that Applies
Classwide Peer Tutoring
Concept-Oriented Reading Instruction
    (CORI)
Content Essentials for Science Destination
    Math
Dyad Reading
Enhanced Wilson Program
Fast ForWord Language
FASTT Math

Fluent Reader
Full Option Science System (FOSS)
Gateways
Hands-on English with Linking Blocks
Help Math
Into English
Intraclient education platform
Language Central for Math
Language Central for Science
Language, Literacy, and Vocabulary
Language Workshop
Math Pathways and Pitfalls
My Reading Coach
Odyssey Math
Open Court
Peer-Assisted Learning Strategies (PALS)
Promoting Science among English
    Language Learners (P-SELLs)
Quality English and Science Teaching
    (QuEST)
Read 180
Success Through Academic Interventions in
    Language & Literacy (SAILL)

Science and Technology for English Language Learners–Achieving Results (STELLAR)
Science interventions
Seeds of Science/Roots of Reading
Snapshots Video Science
Social studies interventions
Spencer Kagan's Cooperative Learning and Active English
SRA Number Words

ST Math
Stevenson Language Skills
Success For All
SuccessMaker
System 44
Systematic and Engaging Early Literacy Instruction
Waterford Early Math and Science Program
Word Generation
Young Reader's Program

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

Campbell Collaboration
Google Education Research
Proquest
WorldCat

EBSCOhost
OVID
PsycINFO

2009 and 2012 Intervention Searches

The primary objective of the intervention search is to identify **all** effectiveness studies conducted for a specific intervention identified in the keyword search, as well as any that the keyword search did not identify. The strategy for the search is as follows:

- If the intervention was reviewed under different WWC topic areas, re-review all references against the protocol for this topic area.

- Conduct standard library searches of the intervention name.[3]

- Scan references to identify possible synonyms for the intervention in the literature and conduct standard library searches of these terms.

- Once potentially eligible studies are identified, request full text and review the reference lists to cross-check search results. Similarly, review relevant literature reviews. Revise search terms as needed.

- Identify seminal researchers associated with the intervention. Conduct full-text searches of the researcher name combined with the intervention name.

- Identify seminal studies of the intervention and conduct searches of the associated citation.

---

[3] A standard library search consists of searching titles and abstracts in each of the databases described above.

- Contact the intervention's developer for a list of known research on the intervention.

All references resulting from these searches will be screened for eligibility.

In addition to the keyword and intervention searches in databases, the review team seeks to identify other relevant studies through the following approaches:

- Public submissions of materials via the WWC website or directly to WWC staff;

- Solicitations made to key researchers by the review team;

- Checking websites summarizing research on programs in science, prior literature reviews, and research syntheses (i.e., using the reference lists of prior reviews and research syntheses to make sure key studies have not been omitted);

- Searches of the websites of all the developers of relevant interventions or practices for any research or implementation reports;

- Searches of the websites of the following think tanks, research centers, and associations:

  Abt Associates
  Alliance for Excellent Education
  American Association for the Advancement of Science
  American Association of Physics Teachers
  American Enterprise Institute
  American Institutes for Research (AIR) Appalachian Education Laboratory (Edvantia)
  Best Evidence Encyclopedia (BEE)
  Broad Foundation (Education)
  Brookings Institution Carnegie Corporation
  Center for Comprehensive School Reform and Improvement
  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 Social Organization of Schools at Johns Hopkins University
  Center on Education Policy
  Center on Instruction
  Chapin Hall Center for Children
  Consortium for Policy Research in Education (CPRE) at the University of Wisconsin-Madison
  Congressional Research Service
  Government Accountability Office
  Harvard Graduate School of Education
  Heritage Foundation
  Hoover Institution
  Horizon Research Inc.
  Inverness Research

Institute for Higher Education Policy
Institute for Public Policy and Social Research (IPPSR)
Johns Hopkins University School of Education
Learning Point Associates
Mathematica Policy Research
MDRC
Mid-Continent Research for Education and Learning
National Association for Bilingual Education (NABE)
National Association of State Boards of Education
National Center on Secondary Education and Transition
National College Access Network
National Dropout Prevention Centers
National Governors' Association
National Science Foundation (NSF)
National Science Resources Center (NSRC)
National Science Teachers' Association (NSTA)
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)
Rand Corporation
Regional Educational Laboratories (RELs)
Southwest Educational Development Laboratory (*SEDL*)
SRI
The Education Resources Institute
The University of California, Los Angeles (UCLA)
Thomas B. Fordham Institute Urban Institute
U.S. Department of Education (includes Institute of Education Sciences)
Wisconsin Center for Education Research (WCER)
WestEd

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.