

TEACHING ACADEMIC CONTENT AND LITERACY TO ENGLISH LEARNERS IN ELEMENTARY AND MIDDLE SCHOOL REVIEW PROTOCOL

This protocol was used to conduct a research review to help inform the recommendations for the What Works Clearinghouse (WWC) updated English learner practice guide, *Teaching Academic Content and Literacy to English Learners in Elementary and Middle School*. The research review involved the following steps:

- The research staff searched the professional literature to identify relevant studies. Additional studies were identified by the expert panel.
- Studies were screened to determine whether they were within the scope of the practice guide.
- Eligible studies were assessed against WWC evidence standards. Studies that met WWC evidence standards and were related to a recommendation within the guide were used to identify the strength of the evidence for each recommendation.¹

This document contains information about: (1) the purpose statement and key definitions that guided the work of the panel and the research team; (2) the procedures that were followed for conducting the literature search; (3) the eligibility criteria used to screen for relevant studies; and (4) technical issues, including attrition and group equivalence, relevant to the review of eligible studies. Please refer to the WWC *Procedures and Standards Handbook (version 2.1)* for additional information.²

PURPOSE STATEMENT

This practice guide aimed to provide teachers with evidence-based recommendations on how to teach English learners effectively. Instruction for English learners involves supporting students as they learn skills and content at the same time as they are learning the English language. The panel examined research on effective instruction and English language development strategies for English learners, emphasizing areas such as academic language (including academic vocabulary), comprehension, and content-area literacy.

KEY DEFINITIONS

English Learners. English learners are students with a primary language other than English who possess a limited range of English speaking, reading, writing, and listening skills.

¹ Some studies that did not meet WWC evidence standards were used to provide examples of practices. (This differs from the procedures for WWC intervention reports, which report findings only for studies meeting WWC evidence standards.)

² What Works Clearinghouse (2011).

This group may include students whose school has identified them as having limited English proficiency at the time of the study or within the preceding two years. Other terms used to describe English learners include Limited English Proficient students, English Language learners, Non-English Speakers, English-as-a-Second-Language learners, Speakers of Other Languages, or Language Minority students.

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

Panel. The panel is the group of individuals who were charged with developing this practice guide, including a panel chair and panelists.

- **Panel Chair** is an expert in his or her field and was responsible for leading the practice guide panelists in developing the practice guide.
- **Panelists** are practice, content, and methodological experts (typically five to seven individuals, including two practitioners) who evaluated research and developed recommendations.

PROCEDURES FOR CONDUCTING THE LITERATURE SEARCH

The literature search involved a keyword search of multiple databases to identify effectiveness studies relevant to teaching English learners.

The review team conducted a keyword search using the Boolean search operators “AND” and “OR” to locate potentially eligible studies. Keywords emanated from the topics the panel was interested in (e.g., writing), as well as the specific search criteria outlined in this protocol (e.g., English learners in grades K-8, residing in the United States or Canada, etc.). Database-specific keywords were located by searching the thesaurus or term library for each database.

The following is a list of the databases that were searched:

- ERIC
- PsycINFO
- Campbell Collaboration
- Dissertation Abstracts
- Academic Search Premier
- Social Sciences Citation Index
- WorldCat

In addition to the keyword search in databases, the review team identified other relevant studies through the following approaches:

- Submissions of materials from the panel
- Solicitations for relevant studies made to key researchers in the field
- Searching for research or implementation reports published on 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 International
- 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

Once these tasks were complete, reference lists located on websites summarizing research on English learners and in prior literature reviews and research syntheses were checked to ensure that relevant studies were not absent from the final list of references to be screened. Relevant studies that were missing were added to the final list of studies to be screened.

References resulting from these approaches were screened and sorted in terms of the recommendations they were used to support.

ELIGIBILITY CRITERIA FOR REVIEWING RELEVANT STUDIES

Studies identified through the literature search were screened for relevance according to the eligibility criteria described in this section.

Populations Included

To be eligible for review, studies had to include English learners. This requirement could be fulfilled in three distinct ways:

- (1) All participants in the study were English learners, *or*
- (2) English learners constituted at least 50% of the sample, *or*
- (3) English learners constituted less than 50% of the sample, but their results were broken out and reported separately from the rest of the sample.

The English learners in the study sample had to be in elementary and middle school (Grades K to 8).³ Studies that contained students in other grades (e.g., pre-K or high school) were not included unless: (a) the study findings disaggregated the results of students in eligible grades or (b) students in eligible grades represented over 50% of the aggregated mixed-age sample. Students in the study had to reside in and attend a school within the United States (including U.S. territories and tribal entities) or Canada (with the stipulation that the students had to be English learners).

Types of Interventions Included

Only studies on replicable interventions were included in the review. For an intervention to be considered replicable, it had to be either a “branded” intervention or an unbranded intervention or practice that met the following conditions:

- (1) The intervention was described in general terms, such as processes and/or skills being targeted, approach to enhancing the processes and/or skills, targeted population, unit of delivery (i.e., whole group, small group, or individual student), and medium/media of delivery (i.e., teacher-led instruction or software),
- (2) The intervention’s duration was described, and
- (3) The characteristics of the individuals who were expected to deliver the intervention were described.

To be eligible for review, interventions had to be at least eight hours in duration.⁴ After-school programs were not eligible for review. The interventions eligible for review could include the following:

Programs. Programs considered for this guide *typically* consisted of educational content. Programs had to be based on text materials, computer software, videotapes, professional development packages for teachers, or any other material base.

Instructional Practices. This review included interventions that focused on educational practices, such as pre-teaching vocabulary words or cooperative learning strategies.

Types of Interventions Excluded

Interventions in which the primary language of instruction was a language other than English were excluded. The following types of studies were also excluded:

³ The panel initially considered developing a guide for students in pre-K to Grade 8. However, given the differences between the instructional needs of pre-K students and those in the primary grades (Grades 1 to 3), the panel decided to limit the scope of the guide to students in Grades K through 8.

⁴ The panel determined the requirement that interventions had to be at least eight hours long. This decision was based on a tool developed by OSEP for evaluating intervention programs for struggling students performing below grade level. However, as the instructional needs of English learners (the student sample targeted in this review) differ from those of struggling students, studies with interventions less than eight hours long, but relevant to the recommendations in the guide, were reviewed at the principal investigator’s discretion.

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

In 2006, the Report of the National Literacy Panel on Language-Minority Children and Youth⁵ examined evaluations that compared the effectiveness of fully immersing English learners in English instruction with teaching them in their native language and transitioning them into English language content instruction when they demonstrated adequate proficiency in English.⁶ The report concluded that such evaluations were problematic because of major design flaws and that they do not provide an empirical basis for determining federal or local policy in regards to which approach to use. Since that time, however, one study that examined this issue was deemed to be of acceptable quality.⁷ This study found no significant differences between English and Spanish being the primary language of academic instruction in the early years of schooling. Thus, the current practice guide panel concurred with the appraisal of the panel for the original English learner practice guide, *Effective Literacy and English Language Instruction for English Learners in the Elementary Grades: A Practice Guide*,⁸ that the focus of the research review would be on studies that explored effectiveness of approaches for teaching students primarily in English, rather than comparisons of native language instruction to English language instruction.

The panel believed that when English learners receive academic instruction in English, some level of strategic native language support is advisable. Therefore, the review team considered studies in which native language was used occasionally (e.g., instructions were given in the native language), while most of the instruction was in English, at the discretion of the principal investigator or panel chair.

Types of Studies Included

Design. Eligible study designs included randomized controlled trials or quasi-experiments with a comparison group.

Publication Date. For the original English learner practice guide, *Effective Literacy and English Language Instruction for English Learners in the Elementary Grades: A Practice Guide*,⁹ the literature search included studies published from 1989 to 2006. The literature search for this current review was extended to September 2012 in the following manner:

- For recommendations from the original English learner practice guide that were carried over to the new guide, studies from the original English learner practice guide (those published between 1989 and 2006) were reviewed again using WWC

⁵ August and Shanahan (2006).

⁶ For example, August and Hakuta (1997); Gersten & Baker (2000).

⁷ Slavin, Madden, Calderon, Chamberlain, and Hennessy (2010).

⁸ Gersten, Baker, Shanahan, Linan-Thompson, Collins, and Scarcella (2007).

⁹ Gersten, Baker, Shanahan, Linan-Thompson, Collins, and Scarcella (2007).

Procedures and Standards Handbook (version 2.1).¹⁰ In addition, the literature search for those topics was extended to include studies from 2006 to September 2012.

- For recommendations that were not mentioned in the original English learner practice guide and did not have any supporting studies from the original English learner practice guide, a literature search was conducted to include studies published from 1989 to September 2012.

Types of Outcomes Included

To be eligible for review, the study had to include at least one relevant student achievement outcome that demonstrated sufficient reliability and face validity. *Reliability* for group-design studies was assessed using the following standards determined by the WWC: internal consistency (minimum of 0.50), temporal stability/test-retest reliability (minimum of 0.40), or inter-rater reliability (minimum of 0.50).¹¹ Although student outcomes were of primary interest for this guide, teacher outcomes, such as observed teaching practice and teacher knowledge, were also reported if relevant to the recommendations.

Relevant outcomes were measures of student achievement, including nationally normed tests, other standardized tests, and researcher-developed measures, in the following six domains: pre-reading, reading, vocabulary, English language development, writing, and content-area acquisition.¹²

1. Pre-reading (for kindergarten only). Pre-reading outcomes included measures of:

- Letter recognition (letter naming),
- Letter sounds,
- Rhyming,
- Beginning sounds, and
- Phonological awareness (e.g., onset rime, phoneme segmentation, blending phonemes).

2. Reading. Reading outcomes included measures of:

- Word reading (including pseudo-word reading),
- Oral or silent reading fluency and/or accuracy in reading connected text,
- Reading comprehension, and
- Measures of overall reading achievement.

¹⁰ What Works Clearinghouse (2011).

¹¹ What Works Clearinghouse (2011).

¹² Content-area experts guided the adaptation of the outcome domains of pre-reading, reading, English language development, and content-area acquisition from the WWC evidence review protocol for the English learners topic area (What Works Clearinghouse, 2013, January). This topic area protocol was intended to be used for evaluating interventions in all areas of K-12 instruction, with a focus on curricular and instructional programs and practices. The protocol's outcome domains were determined to be appropriate for the topics the panel chose to pursue for the current practice guide review and were adapted to fit the goals and scope of the current review.

3. Vocabulary (including academic vocabulary¹³).¹⁴ Vocabulary included measures of:

- Receptive vocabulary (oral and written), and
- Expressive vocabulary.

4. English language development. English language development included measures of:

- Listening comprehension,
- Grammar/syntax, and
- Other linguistic features of the English language.

5. Writing.¹⁵ Writing outcomes included measures of:

- Overall writing quality,
- Writing output,
- Mechanics,
- Organization, and
- Sentence structure.

6. Content-area acquisition. Content-area acquisition included measures of:

- Science,
- Social studies, and
- Math.

Overalignment of outcomes. To be eligible for review, the outcome measures must not have been overaligned with either one of the experimental conditions (intervention or comparison), as this could have biased a study's results. For instance, an outcome measure could have been considered to be overaligned with the intervention condition if it assessed achievement using some of the same materials included in the intervention (such as specific questions). In such situations, the intervention group could have had an advantage over the comparison group, and the size of the intervention's measured effects could be incorrect.¹⁶

¹³ Academic vocabulary represents a set of words that are used in academic classrooms and text much more often than in everyday social and informal settings. Academic vocabulary words include both general academic words and domain-specific words.

¹⁴ At the time that this protocol was written, researchers, policymakers, and practitioners focused strongly on improving "academic vocabulary." Although academic vocabulary outcomes could be part of both reading and English language development domains, the panel decided to form a separate domain for vocabulary outcomes, to reflect the interest in the field and to inform the panel's recommendations.

¹⁵ Writing outcomes used in this guide were based on the outcome categories used in a recently published WWC practice guide on writing. See Graham et al. (2012).

¹⁶ In some of the studies that were reviewed, the vocabulary outcome measures included words that were part of the intervention condition, and it was not clear whether the same words were also part of the control condition. Consequently, issues of bias and overalignment were of concern with vocabulary outcome measures. However, for this review, such vocabulary measures were not considered as being overaligned by the panel, for the following

STATISTICAL AND TECHNICAL ISSUES

Each study that met the screening criteria was reviewed by WWC-certified reviewers, who completed the Study Review Guide (SRG) for group designs, taking care to provide complete, concise, and correct information. The reviewers determined each study's evidence of causal validity according to evidence standards delineated in the *WWC Procedures and Standards Handbook (version 2.1)*.¹⁷

Reviewers used the SRG to assign each study one of three possible ratings:

1. **Meets WWC Evidence Standards without Reservations.** This rating was used for randomized controlled trials (RCTs) that provided the strongest evidence of causal validity and did not have problems with randomization or sample attrition.
2. **Meets WWC Evidence Standards with Reservations.** This rating was used for strong quasi-experimental design (QED) studies that had valid comparison groups and met other WWC evidence standards, as well as some RCTs that had problems with randomization, attrition, or disruption.
3. **Does Not Meet WWC Evidence Standards.** This rating was used for studies that provided insufficient evidence of causal validity.

If a study received a rating of either “Meets Standards without Reservations” or “Meets Standards with Reservations,” then the study was reviewed by a second reviewer. Information from both reviewers was combined to create a Master SRG, and a reconciliation meeting was scheduled to discuss discrepancies.

Additional Information Collected on the SRG:

A thorough description of the practices or curricula each treatment and/or comparison group received was required to assist the panelists in successfully incorporating the reviewed evidence into the guide. However, no study failed to meet standards due to an incomplete description. Reviewers included information on the following, if available:

- Student native language proficiency
- Student English language proficiency at pre-test
- Other academic pre-test measures
- Length of intervention
- Intensity of intervention (i.e., number of hours of instruction)

(continued)

reasons. The panel noted that if students were given measures with words not taught in the program, then those measures were not likely to capture the effectiveness of the vocabulary program. The panel also felt that the issue of alignment was somewhat mitigated as students were required to learn and retain knowledge of the words for a minimum of two months prior to being tested.

¹⁷ What Works Clearinghouse (2011).

- Description of the intervention (program or practice)
- Teacher training in intervention strategy
- Context of instruction (e.g., special education, ESL class, regular classroom, dual-immersion classroom)
- Material such as texts, videotapes, software, or other classroom materials
- English learner definition used by the school or researcher
- Eligible outcome measures

Attrition Levels in RCTs

According to the *Procedures and Standards Handbook (version 2.1)*,¹⁸ both the overall sample attrition rate and the differential in sample attrition between the intervention and comparison groups should be examined, as both contribute to the potential bias of an intervention's estimated effect. The WWC has established conservative and liberal standards for acceptable attrition levels. Conservative standards are typically applied in cases where the principal investigator has reason to believe that much of the attrition is endogenous to the intervention reviewed (e.g., high school students choosing whether or not to participate in an after-school program). Liberal standards are typically applied in cases where the principal investigator has reason to believe that much of the attrition is exogenous to the intervention reviewed (e.g., in cases where young children's movement in and out of school districts was 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.

For this review, a liberal attrition standard was used, reflecting the assumption that most attrition in studies of English learners was due to factors that were not strongly related to treatment status, such as parent mobility and absences on the days when assessments were conducted. Table 1 presents the maximum difference in the attrition rate for the treatment and comparison groups that was acceptable for a given level of overall sample attrition. The empirical basis for these thresholds is described in Appendix A of the *Procedures and Standards Handbook (version 2.1)*.¹⁹

¹⁸ What Works Clearinghouse (2011).

¹⁹ What Works Clearinghouse (2011).

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	-

Studies based on cluster random assignment designs had to meet attrition standards for both the study sample units that were assigned to treatment and control 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 included only sample members in the clusters that remained in the study sample.

RCTs with combinations of overall and differential attrition rates that exceeded the applicable threshold, based on the applicable standard, had to demonstrate baseline equivalence of the analysis sample, or, if nonequivalence fell 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 next section on Baseline Equivalence for more details.

Baseline Equivalence

RCTs with high attrition and all QEDs had to demonstrate baseline (i.e., 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*. Baseline equivalence was examined on pre-test measures, either a pre-test of an acceptable outcome measure or other baseline measures that were expected to be highly correlated with the outcome measures. Within each outcome domain, there were established rules for acceptable baseline measures. These included:

- (1) Only measures from a particular domain could be used in establishing baseline equivalence for that domain. For example, for reading, only reading-related outcomes were used to establish baseline equivalence; for science, only science-related outcomes were used, and so on. Thus, a reading-related outcome was not used to establish equivalence in the science domain.
- (2) In the reading domain, in addition to the outcomes listed on page 7, measures such as phonemic awareness and alphabetic knowledge in English were also used to establish equivalence for K through Grade 1 studies.
- (3) If a domain included more than one applicable measure, then equivalence had to be established on all those measures.

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

If the differences were greater than 5% but less than or equal to 25% of the pooled standard deviation in the sample, then the study received a rating of *meets WWC evidence standards with reservations* as long as the study findings were based on analytic models that controlled for the individual-level baseline characteristic(s) on which the groups differed. Otherwise, the study was rated *does not meet WWC evidence standards*.

If the baseline differences were greater than 25% of the pooled standard deviation for *any* of the measures (in *any* of the domains), then the outcome domain did not meet the baseline equivalence standard, regardless of whether or not the impacts were estimated using models that controlled for baseline characteristics. Therefore, the domain was rated *did not meet WWC evidence standards*.

Statistical and Analytical Issues

Based on the *WWC Procedures and Standards Handbook (version 2.1)*,²⁰ for this review, RCTs with low attrition were not required to use statistical controls in the analysis, although statistical adjustment for well-implemented RCTs was permissible as it could help generate more precise effect-size estimates. For RCTs with high attrition and baseline differences less than .05 of a standard deviation and QEDs with baseline differences less than .05 of a standard deviation, effect-size estimates were adjusted for differences in pre-intervention characteristics at baseline (if available) using a difference-in-differences method if the authors did not adjust for pre-test. Beyond the pre-intervention characteristics required by the equivalence standard, statistical adjustment was also made for other measures in the analysis, though not required.

The preference was to calculate effect sizes using post-intervention means adjusted for the pre-intervention measure. If a study reported both unadjusted and adjusted post-intervention means, then adjusted means and the unadjusted standard deviations were used for effect size calculations. If adjusted post-intervention means were not reported, then group means were calculated using a difference-in-differences approach.

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

When a misaligned analysis was reported (i.e., the unit of analysis in the study was not the same as the unit of assignment), the effect sizes were computed by incorporating a statistical adjustment for clustering. The default intraclass correlation used for the English learner review was 0.20 for all domains. For an explanation about the clustering correction, see Appendix C of the *Procedures and Standards Handbook (version 2.1)*.²¹

When multiple comparisons were made (i.e., multiple outcome measures were assessed within an outcome domain in one study) and not accounted for by the authors, the effect's reported statistical significance was adjusted using the Benjamini-Hochberg correction. See Appendix D of the *Procedures and Standards Handbook (version 2.1)* for the formulas used to adjust for multiple comparisons.²² (All standards applied to overall findings as well as analyses of sub-samples.)

²⁰ What Works Clearinghouse (2011).

²¹ What Works Clearinghouse (2011).

²² What Works Clearinghouse (2011).

References

- August, D., & Shanahan, T. (Eds.). (2006). *Developing Literacy in Second-language Learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- August, D. & Hakuta, K. (1997). *Improving schooling for language-minority children: A research agenda*. Washington, DC: The 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.
- Graham, S., Bollinger, A., Booth Olson, C., D'Aoust, C., MacArthur, C., McCutchen, D., & Olinghouse, N. (2012). *Teaching elementary school students to be effective writers: A practice guide* (NCEE 2012-4058). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications_reviews.aspx#pubsearch/
- Slavin, R. E., Madden, N., Calderon, M., Chamberlain, A., & Hennessy, M. (2010). *Reading and language outcomes of a five-year randomized evaluation of transitional bilingual education*. Baltimore, MD: Johns Hopkins University.
- What Works Clearinghouse (2011). *Procedures and standards handbook (version 2.1)*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handbook.pdf
- What Works Clearinghouse. (2013, January). Evidence Review Protocol for Interventions for English Language Learners (Version 2.2). Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_ell_protocol_v2.2.pdf