REVIEW PROTOCOL FOR
BEGINNING READING INTERVENTIONS
VERSION 3.0 (SEPTEMBER 25, 2014)

This review-specific protocol guides the review of research that informs the What Works Clearinghouse (WWC) intervention reports in the Beginning Reading topic area. The review-specific protocol is used in conjunction with the *WWC Procedures and Standards Handbook (version 3.0)*.

PURPOSE STATEMENT

This review focuses on beginning reading interventions designed for use with students in grades K–3 (or ages 5–8), with a primary focus on increasing literacy skills.

The following research questions guide this review:

- Among interventions intended to provide literacy instruction, which ones improve literacy skills (alphabetics, reading fluency, comprehension, and general reading achievement) among students in grades K–3?

- Are some interventions more effective than others for certain types of literacy skills?

- Are some interventions more effective for certain types of students, particularly students who have historically lagged behind in reading and/or literacy achievement?

KEY DEFINITIONS

Beginning reader. Beginning readers are defined as students in grades K–3 in classes with a primary focus on improving literacy skills (such as English/Language Arts classes).

Reading intervention. In this review, a reading intervention is defined as a replicable (i.e., can be reproduced) instructional program that is delivered to students, clearly delineates literacy learning goals for students, and is designed to directly affect student reading achievement.

PROCEDURES FOR CONDUCTING THE LITERATURE SEARCH

The *WWC Procedures and Standards Handbook, version 3.0,* discusses the procedures for conducting a literature search in Section II: Developing the Review Protocol and Identifying Relevant Literature (p. 4) and in Appendix B: Policies for Searching and Prioritizing Studies for Review.

Search Terms

The following table presents the search terms by category.

Category	Search Terms	
Study Design	• ABAB design	• Pretest
	• Affect*	• Pre-test
	• Assignment	• QED
	• Causal	• QES
	• Comparison group	• Quasi-experimental
	• Control*	• Quasiexperimental
	• Counterfactual	• Random*
	• Effect*	• RCT
	• Efficacy	• RDD
	• Evaluation*	• Regression discontinuity
	• Experiment*	• Simultaneous treatment
	• Impact*	• SCD
	• Matched group	• Single-case
	• Meta analysis	• Single subject
	• Meta-analysis	• Treatment
	• Posttest	• Reversal design
	• Post-test	• Withdrawal design
Intervention (Broadly)	• Approach	• Program
	• Curricul*	• Remedial
	• Educational therapy	• School*
	• Homework	• Strategy
	• Improvement	• Success*
	• Instruct*	• Teach*
	• Practice	• Treatment
Population	• Adolescent*	• Grade 1
	• Elementary school	• Grade 2
	• Kindergarten	• Grade 3
	• First grade	• K–12
	• Second grade	• Student*
	• Third grade	• Summer school

Category	Search Terms	
Outcomes	• Alphabetics • Aural learning • Comprehension • Fluency • Language • Letter identification • Lexicography • Literacy • Phonemic • Phonetics	• Phonics • Phonological • Print awareness • Print knowledge • Readability • Reading • Verbal development • Vocabulary • Vocalization • Word recognition

Additional Sources

This review searched the electronic databases listed in the *WWC Procedures and Standards Handbook*, Appendix B.

In addition to those listed in the *WWC Procedures and Standards Handbook*, Appendix B, this review searched the following websites:

- American Evaluation Association (AEA)
- American Federation of Teachers (AFT)
- Association for Public Policy Analysis and Management (APPAM)
- 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 on Education Policy
- Center on Instruction
- Consortium for Policy Research in Education (CPRE)
- ICF International
- National Education Association (NEA)
- National Institute of Child Health and Human Development (NICHD)
- National Reading Panel
- NBER Working Papers
- Pacific Resources for Education and Learning (PREL)
- Public Policy Research Institute at Texas A&M University
- Southwest Educational Development Laboratory (SDEL)
- Westat

ELIGIBILITY CRITERIA

Eligible Populations

In this review, the following populations are of interest:

- *Grade range.* The Beginning Reading topic area will review studies of interventions administered to students in grades K–3. The primary criterion for determining if a study is eligible for review is the grade range. Studies with students in grades K–3 are eligible for review. Studies that include students who are older than 8 and younger than 5 are included as long as the students are in grades K–3. If authors do not provide the grade for study students, the review will use the age range of 5–8 to determine if the study is eligible.

- *Location.* The intervention must be provided to students in an academic setting, including elementary schools, secondary schools, summer school programs, or home-schooling programs.

- *Overlap with other WWC topic areas.* Studies of students who are typically developing, as well as those at risk for reading difficulties, including students with learning disabilities and low-income minority students, are considered for Beginning Reading reviews. To be included in the review, at least 50% of the students in the study must be general education students (which excludes students classified as English learners [ELs] or receiving special education services, as these students are included in other WWC topic area reviews).

  Studies of reading interventions involving students in grades 4–12 are reviewed under the WWC Adolescent Literacy topic area. For studies that include samples of students that span both the Beginning Reading and Adolescent Literacy topic areas and cannot be disaggregated by grade level, the Beginning Reading topic area will review any studies that include fourth-grade students and lower (for example, a combined sample of students from grades K–4). Any studies in which students receive the intervention in grade 5 (or higher) will be reviewed by the Adolescent Literacy topic area (for example, a study in which receipt of the intervention spans grades 2–5).

Potential subgroups of interest for this review include:

- Characteristics of students:

  o Age

  o Gender

  o Race/ethnicity

  o Socioeconomic status (SES)

- o Pre-intervention reading achievement levels (i.e, poor readers or underachievers)

- Characteristics of intervention settings:

  - o Location of the schools involved (e.g., urban, suburban, rural)

  - o School type (public, private, religious)

  - o Average class size (small, medium, large)

Eligible Interventions

Only reading interventions that are replicable (i.e., can be reproduced) 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) (e.g., strategies, activities, and materials), unit of delivery of the intervention (for example, whole group, individual), medium/media of delivery (for example, teacher-led instruction or software), and targeted population;

- Intervention duration and intensity; and

- Description of individuals delivering or administering the intervention.

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

- *Curricula.* The review includes (a) comprehensive, non-textbook-based curricula intended to serve as a school's primary literacy instruction program; (b) supplemental programs that are intended to enhance a whole-school or whole-classroom literacy curricula; or (c) programs aimed at struggling readers and students who read behind their grade level, such as remedial curricula. Professional development for the particular curricula is considered a part of the curricula.

- *Practices or strategies.* The review includes (a) classroom practices or strategies intended to address a specific literacy-related skill (such as semantic mapping, vocabulary instruction, questioning, summarizing) or (b) school-wide policies (such as a school-wide literacy initiative). As with curricula, any professional development provided to help implement the practice or strategy is considered part of the practice or strategy.

- *Products.* The review includes reading or literacy textbooks intended for whole-school or whole-classroom use and software designed to improve literacy skills.

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 (e.g., instructions/guidance on the implementation of the intervention) or sells or distributes the intervention.

- Trademark or copyright.

Eligible Research

The *WWC Procedures and Standards Handbook* discusses the types of research reviewed by the WWC in Section II: Developing the Review Protocol and Identifying Relevant Literature (p. 4). Additionally, in this review, the following additional parameters define the scope of research studies to be included:

- *Topic.* The intervention must focus on the effects of a reading or literacy intervention on one or more measures of reading or literacy, including alphabetics, reading fluency, comprehension, or general literacy achievement.

- *Time frame.* The study must have been released in 1983 or later and be obtained by the WWC for review prior to the drafting of the intervention report.

- *Sample.* The study sample must meet the requirements specified above in the "Eligible Populations" section. Outcomes can be measured later (e.g., when the sample is older).

- *Language.* The study must be available in English to be included in the review. Also, studies examining reading competencies in other languages will not be included in the review.

- *Location.* The study must include students in the United States, its territories or tribal entities, or in a country that is sufficiently similar to the United States that the study could be replicated in the United States (e.g., in which English is the societal language). DoDEA schools on overseas US military bases can also be considered an acceptable setting.

Eligible Outcomes

This review includes outcomes in the following domains:

Alphabetics domain. The alphabetics domain includes outcomes measuring phonemic awareness, phonological awareness, letter identification, print awareness, and phonics. Each outcome is described in more detail below.

- *Phonemic awareness outcomes.* Phonemic awareness (or phoneme awareness) refers to the understanding that the sounds of spoken language—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").

- ***Phonological awareness outcomes.*** Phonological awareness is a more encompassing term than phoneme/phonemic awareness. It refers to phoneme awareness and to awareness of larger spoken units such as syllables and rhyming words. Tasks 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 (e.g., "what is cowboy without cow?").

- ***Letter identification outcomes.*** Letter identification refers to knowledge of the names of the letters of the alphabet.

- ***Print awareness outcomes.*** Print awareness refers to 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 (e.g., author, title, front/back).

- ***Phonics outcomes.*** Phonics[1] refers to (a) the knowledge 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) the ability to associate letters and letter combinations with sounds and blending them into syllables and words, and (c) the understanding that this information can be used to read or decode words. Spelling is included in the review 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.*

Reading fluency domain. The reading fluency domain includes outcomes measuring fluency, or the ability to read text accurately, automatically, and with expression (including appropriate pausing, response to punctuation, and so on), while extracting meaning from it.

Comprehension domain. The comprehension domain includes outcomes measuring vocabulary development and reading comprehension. Each outcome is described in more detail below.

---

[1] "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 the Beginning Reading review protocol, we use the term *phonics* as defined above, not as an instructional approach.

- ***Vocabulary development outcomes.*** Vocabulary development refers to the development of knowledge about the meanings and uses of words. The development of receptive vocabulary (words understood) and expressive vocabulary (words used) is critical for reading comprehension.

- ***Reading comprehension outcomes.*** Reading comprehension refers to the understanding of 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. Struggling readers may have difficulty with any of these components of reading or with multiple components. Reading comprehension outcomes may include tests of students' comprehension of passages from various content areas. For example, a test assessing students' comprehension of a social studies passage would be an acceptable outcome.[2]

General reading achievement domain.[3] Outcomes that fall in the general reading achievement domain combine separate measures of two or more of the previous domains (alphabetics, reading fluency, and comprehension) by providing some type of summary score across domains, such as a "total reading score" on a standardized reading test.

EVIDENCE STANDARDS

Eligible studies are assessed against WWC evidence standards, as described in the *WWC Procedures and Standards Handbook,* Section III: Screening and Reviewing Studies (pp. 8–21).

Sample Attrition

The *WWC Procedures and Standards Handbook* discusses the sample attrition standards used by the WWC in Section III: Subsection B.2 Sample Attrition: Is the combination of overall and differential attrition high? (pp. 11–15).

This review uses the ***liberal*** boundary for attrition. This boundary was based on the assumption that most attrition in studies of Beginning Reading 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. The *WWC Procedures and Standards Handbook* contains a figure illustrating the attrition boundary and an associated table with

---

[2] These comprehension tests assess the student's ability to read a passage and answer questions based on the material in the text and require no prior knowledge of the material in order to complete the assessment successfully. Content area knowledge tests that assess an individual's preexisting understanding of the facts, theories, and other related materials in that area (e.g., science, social studies, mathematics) are not eligible outcomes.

[3] Although writing outcomes are outside the scope of the Beginning Reading review, Language Arts outcomes, such as English language conventions (e.g., sentence structure, grammar, etc.), are eligible for review.

attrition levels that define high and low attrition. Based on the choice of the boundary, the study review guide calculates attrition and whether it is high or low.

Baseline Equivalence

If the study design is a randomized controlled trial or regression discontinuity design with high levels of attrition or a quasi-experimental design, 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. The *WWC Procedures and Standards Handbook* discusses how authors must demonstrate baseline equivalence in Section III: Subsection B.3 Baseline Equivalence: Is equivalence established at baseline for the groups in the analytic sample? (pp. 15–16).

Baseline equivalence must be demonstrated for the intervention and comparison groups in the analytic sample on one of the following pre-intervention (or baseline) characteristics:

- A pre-intervention measure of the outcome used in the analysis; or

- If a pre-intervention measure of the outcome used in the analysis is not available, a pre-intervention measure of an outcome from any of the four outcome domains detailed in the "Eligible Outcomes" section above can be used. For example, a pretest from an alphabetics domain can be used to establish baseline equivalence for the reading fluency outcome when a reading fluency pretest is unavailable.

If a study has a pre-intervention difference greater than 0.25 SD for *any* outcome in the four literacy domains, the entire study is rated *does not meet WWC group design standards*.

This review requires that, in a domain that requires statistical adjustments, the adjustment is made for all outcomes in the domain. For example, if A, B, and C are available as pre- and post-intervention measures within the same domain, and the pre-intervention difference in B requires statistical adjustment, each analysis of an outcome (A, B, or C) must adjust for B.

Review team leadership should be notified if a study has a baseline difference greater than 0.25 SD in any of the following characteristics, since it could be evidence that the populations were drawn from very different settings and that the intervention and comparison groups are not sufficiently comparable for the purposes of this review:

- Gender/sex
- Race/ethnicity
- Percentage of ELs
- Percentage of students receiving special education and related services
- Socioeconomic status (SES)
- School location (urban, rural, suburban)
- Average class size (small, medium, large)

9

The provision of such information, however, is not a requirement of the review.

Outcomes

In this review, the requirements for outcome measures differ from guidance in the *WWC Procedures and Standards Handbook* (in Section III: Subsection B.4 Outcome Eligibility and Reporting [pp. 16–19]) in the following way: This review follows more stringent guidance with respect to internal consistency (such as Cronbach's alpha). This review requires a minimum of 0.60 for internal consistency. This review prioritizes end-of-intervention (immediate posttest) outcomes that are measured after full implementation of the intervention (i.e., outcomes administered after the third year of a 3-year intervention is completed).

In this review, the general guidance regarding composite and subscale scores, subgroup findings, categorical ordinal measures, and estimated effects using imputed data are followed.

Statistical Adjustments

The *WWC Procedures and Standards Handbook* discusses the types of adjustments made by the WWC in Section IV: Subsection B. Statistical Significance of Findings (p. 24).

Other Study Designs

Studies that use regression discontinuity or single-case designs are eligible for review using the appropriate pilot standards.

The *WWC Procedures and Standards Handbook* discusses the pilot standards for reviewing regression discontinuity design studies in Appendix D.

The *WWC Procedures and Standards Handbook* discusses the pilot standards for reviewing single-case design studies in Appendix E.