REVIEW PROTOCOL FOR
FOUNDATIONAL READING PRACTICE GUIDE
VERSION 3.0 (NOVEMBER, 2014)

This review-specific protocol guides the review of research that informs the What Works Clearinghouse (WWC) practice guide "Foundational Reading." The review-specific protocol is used in conjunction with the *WWC Procedures and Standards Handbook (version 3.0)*.

PURPOSE STATEMENT

This review focuses on interventions or practices designed to help children in kindergarten through third grade learn to develop skills that are considered foundational to developing solid reading abilities. The following research question guides this review: "Which instructional practices improve beginning readers' foundational reading skills?" Specific recommendations in the guide will center on questions like:

- Should phonemic awareness always be introduced in the context of letters?
- How can teachers boost students' oral language skills so as to improve their reading for understanding?
- What words and strategies should I include in teaching vocabulary?
- Does explicit, systematic phonics instruction mean skills are to be taught in isolation?
- Can foundational reading skills be taught embedded in meaningful context?
- What kind of text supports the teaching of beginning reading skills?

The guide will account for the teaching of foundational reading skills at different elementary grade levels by describing how to adapt practices for students in different grades and by focusing some recommendations on the teaching of foundational reading skills at specific grade levels.

PROCEDURES FOR CONDUCTING THE LITERATURE SEARCH

The *WWC Procedures and Standards Handbook* 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	• Control group* • Comparison group* • Matched group* • Treatment* • Random* • Assign* • Baseline • Experiment* • Evaluat* • Impact* • Effectiveness • Causal* • Post*test* • Pre*test* • Randomized controlled trial	• RCT • Quasi*experiment* • QED • Regression discontinuity • Changing criterion • Intrasubject replication • Multiple baseline • Multi*element • Single case • Single subject • ABAB • Alternating treatment • Simultaneous treatment • Reversal design • Withdrawal design
Topic	• Reading • Literacy • Phonem* • Phonologic* • Rhym* • Decod* • Endoc*	• Oral language • Oral language comprehen* • Syntax • Morphology • Alphabetic principle • Spell* • Writ*

	• Letter names • Letter sounds • Orthograph* knowledge • Orthograph* aware* • Oral language fluen* • Print* • Text* • Oral reading fluen*	• Letter knowledge • PGC[1] • Phonem*[1] • Phonics[1] • Word[2]
Intervention	• Interven* • Curricul* • Program* • Strateg* • Instruct* • Teach*	• Train* • Approach* • Monitor* • Treat* • Self-regulat* • Transfer
Population	• K-3* • Kindergart * • Elementary • First*grade* • Second*grade* • Third*grade* • Age*5 • Age*6	• Age*7 • Age*8 • Age*9 • 5 year* old • 6 year* old • 7 year* old • 8 year* old • 9 year* old
Outcomes	• Achieve* • Improve* • Instructional effectiveness • Outcome* • Effect* • Develop*	• Skill* • Assess* • Test* • Progress* • Acqui*

[1] Term must be found within 10 words of "reading" or "literacy."

[2] Term must be found within 10 words of "analysis" or "recognition" or "identification" or "reading."

Additional Sources

In addition to those databases listed in the *WWC Procedures and Standards Handbook*, Appendix B,[3] this review will also search the EJS E-Journals electronic database.

The review team will search the WWC database of previously reviewed studies to identify studies that have met standards in prior reviews.

The review team will solicit study recommendations of publicly available studies from panel members.

ELIGIBILITY CRITERIA

Eligible Populations

In this review, the following populations are of interest:

- ***Grade range.*** Students in kindergarten through third grade (ages 5 years 0 months, through 9 years 11 months), or in any subset of these grades. Studies that contain students in other grades will not be included unless (a) study reports disaggregated results for students in eligible grades, or (b) students in eligible grades represent the majority of the aggregated mixed-age sample. If the study does not make explicit the number of students in each grade, a study will be included if 50% or more of the grades included in the sample falls within the eligible grade range. If the study provides only the mean age of the sample without any grade information, the mean age must be larger than 5 years 0 months but smaller than 9.5 years (9 years 6 months).
- ***Location.*** Studies can occur outside the United States, but practices and interventions must be conducted in English with primarily English-speaking students.
- ***Language and ability-based subgroups.*** To be included in the review, at least 50% of the students in the study must be general education students and native speaking students (or there must be a subgroup analysis for these students). Students classified as English learners (ELs) or receiving special education services have distinct needs from general education or native speaking students, and are the focus of other practice guides. Further, the guide recommendations will focus on research-supported practices administered in English. The panel will consider:
    - Studies of students that speak English and another language

[3] The search did not include two electronic databases listed in Appendix B: SAGE or ProQuest.

- Studies with primarily students who have limited English proficiency, although reviews will focus on results that disaggregate the English-speakers from other students if such disaggregated results are available
- Studies of at-risk students, students receiving remedial instruction, or other students who may be receiving extra assistance but who do not have an identified disability. For studies that include both general education students and students at-risk, receiving remedial instruction, or other extra assistance, the review will focus on disaggregated results if available.

Eligible Interventions

The review will consider studies of comprehensive or supplemental curricula or replicable instructional strategies for teaching foundational reading skills to students in kindergarten through 3rd grade. These may include strategies or curricula used by teachers in general education classrooms, those used by reading coaches in the school, or those for use by paraprofessional educators, tutors, or parents. The following characteristics of an intervention must be known 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
- Description of individuals delivering or administering the intervention

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

- ***Curricula***. A curriculum is a set of activities, materials, and/or guidance for working with students in classrooms. A curriculum has a clearly identified name, includes a write-up/description, and can be replicated by others based on written guidance, staff training, or technical assistance (for example, *The Me Book*). A curriculum may be (1) intended as the primary instructional tool designed to meet children's learning needs in multiple areas; or (2) designed to supplement the classroom material with differentiated instruction or meet children's learning needs in specific areas. Both types of curricula will be included in this review.
- ***Practices.*** A practice is a named approach to promoting children's development that educators implement by interacting with children and materials in classrooms. The review will include named practices that are clearly described, commonly

understood, and used in published works by more than one investigator or team of investigators. Several terms may be used in the literature to refer to the same practice. A named practice may also refer to an array of specific procedures.

The review excludes (1) practices related to professional development, teacher preparation, and textbook design issues, and (2) other interventions not appropriate for a teacher's practice guide on reading, such as comprehensive school reform. The guide also will not focus on the teaching of reading to English language learners or to students with learning or other disabilities (though the panel will consider studies of reading interventions implemented to groups of students that include, but are not wholly composed of, individuals from these populations when issuing their recommendations).

Both "branded" and "non-branded" interventions will be reviewed. Branded interventions are commercial or published practices 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 recommendations in the practice guide will focus on instructional strategies that improve the foundational reading skills of children in kindergarten through grade three.
- ***Time frame*****.** The study must have been published between 2000 and November 2014; earlier or later work will be reviewed if suggested by a panelist.
- ***Sample.*** The study sample must meet the requirements described in the "Eligible Populations" section above.
- ***Language.*** The study must be available in English to be included in the review. Studies examining instruction in other languages will not be included in the review.
- ***Location.*** Studies can occur outside the United States, but interventions must be administered in English.

- ***Publication.*** Conference papers and dissertations are ineligible, unless requested by a panelist.

Eligible Outcomes

The panel is primarily concerned with interventions or practices designed to help children learn to develop those skills that are considered foundational to developing solid reading abilities. This review includes outcomes in the following domains:

- ***Encoding.*** Includes measures of spelling, invented spelling (attempts to spell), and orthographic knowledge (the knowledge of a language's writing system).
- ***General Achievement.*** Includes measures that are composite or aggregated scores using subtests that cross multiple domains.
- ***Letter names/letter sounds.*** Assess a student's knowledge of the names and sounds of the letters of the alphabet. These tasks involve linking letters to print.
- ***Listening Comprehension.*** Refers to the ability to follow, process, and understand spoken language, including comprehension of informational and narrative text.
- ***Morphology.*** Refers to the knowledge of meaningful word parts in a language (typically the knowledge of prefixes, suffixes, and/or roots and base words).
- ***Oral reading accuracy.*** Includes tasks involving passages that students read aloud. Oral reading accuracy refers to the ability to read a given passage of text aloud accurately (i.e., reading the words correctly) but does not address reading rate. In some tests, results are reported in the form of the percentage of words read accurately; in other tests, students read several text selections of increasing difficulty, and the score represents the highest text level a student can read at a predetermined level of accuracy (e.g., 90 percent accuracy).
- ***Oral reading fluency.*** Refers to fluent reading in connected text. Oral reading fluency is the ability to read a passage of text aloud accurately, at an appropriate rate, and with expression (i.e., with appropriate prosody, including appropriate pausing and oral interpretation of the text). ORF measures used by researchers are typically assessments of rate and accuracy, with results reported as the number of words read correctly in a specified amount of time (e.g., words correct per minute). In this practice

guide, such assessments will be referred to as measures of rate and accuracy. Prosody is typically rated in terms of reading expression and phrasing. In this practice guide, such rating scales will be referred to as measures of prosody.

- ***Phonology.*** Refers to the sound structure of language. Phonology tasks are auditory/oral tasks that focus on students' ability to articulate the sounds without involving letter or word knowledge. Once letters are involved (such as linking sounds to letters), the task is no longer considered phonology, but would belong in either letter names/sounds, word reading or encoding (depending on the nature of the task). This domain includes phonological awareness and phonemic awareness. Phonology tasks include blending onsets and rimes or individual phonemes into words (e.g., /s/ /un/ into sun) or segmenting words into their onsets and rimes or into their individual phonemes (e.g., "Say the sounds in sun."—"/s/ /u/ /n/."). (Onset refers to the part of the syllable that precedes the vowel; rime refers to the part of the syllable that contains the vowel and any consonants following it.) Measures of elision, which require students to delete specific sounds from spoken words (e.g., "Say play without the /p/." —"Lay."), are also phonology tasks. 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. Phonological awareness is a more encompassing term than phonemic awareness. It refers to phoneme awareness and to awareness of larger spoken units such as syllables and rhyming words.

- ***Reading comprehension.*** Refers to the understanding of the meaning of a passage and the context in which the words occur. Reading comprehension depends on various underlying components including decoding (the ability to translate words into speech), knowledge of word meanings, fluency, and the ability to understand and interpret spoken language. We include in this outcome domain several ways of measuring comprehension, including oral or written retelling, reading text and answering questions about it, providing missing words to complete sentences or passages so that they make sense (i.e., cloze tasks), selecting words from a list to complete a passage so that it makes sense (e.g., maze, which is a multiple-choice cloze task). We also include in this category tests of silent-reading efficiency, which are timed measures of silent reading. These tests include maze tasks and sentence-verification tasks, in which students read sentences silently and are asked to verify at the end of each sentence whether it is true or false.

- ***Syntax.*** Refers to the formation of sentences and the associated grammatical rules. Syntax includes outcomes related to understanding sentence structure and recalling sentences.

- ***Vocabulary.*** Refers to the development of knowledge about the meanings, uses, and pronunciation of words, and includes receptive vocabulary (words understood) and expressive vocabulary (words used).

- ***Word reading.*** Includes decoding, word recognition, word identification, and word analysis. Word reading tasks are typically untimed lists of words (and nonwords) that students read aloud and the score is the number of words read correctly. If the task is timed, such as the TOWRE task, then that task is called timed word reading, timed decoding, or word reading efficiency and the score is the number of words read correctly in X time (e.g., 45 sec.). There are also single word decoding tasks where students read words (or nonwords) in a list silently and select the word pronounced by an examiner (or computer).

Measures testing rapid automatized naming (RAN) skills are not eligible for the Foundational Reading Skills evidence review. These measures include timed tests that ask students to repeatedly name a small set of letters, objects, colors, or shapes, as quickly as possible.

Eligible outcomes must be experimenter-designed measures that demonstrate face validity and reliability in measuring a student's skills in learning to read OR be standardized tests. When reliability information is not available, the panel chair and panelists will be consulted to determine the eligibility of the outcome measure. Review team leadership will also consult with the panel chair and panelists as needed to confirm the relevance of study outcome measures to the domains.

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 foundational reading interventions is due to factors that are not strongly related to treatment status, such as parent mobility and absences on the days that assessments are conducted. The *WWC Procedures and Standards Handbook* contains a figure illustrating the attrition boundary and an associated table with 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 and 16).

Equivalence must be established in the domain of the outcome measure. If baseline differences exceed 0.25 standard deviations for any of the measures within a domain, the study will not meet evidence standards within this domain.

For this review, in those cases where a pretest from the same domain is not available or developmentally appropriate, a pretest in one of the following domains can serve as a proxy to make a determination of equivalence: letter names and sounds, phonology, and/or word reading. If a pretest in more than one of these three proxy domains is present, equivalence will be determined using all available proxy pretests, according to the following rules:

- If any of the available proxy pretests has a difference that exceeds 0.25 SD, the domain using the proxy test will be determined to be not equivalent.

- If any of the available proxy pretests has a difference less than 0.05 SD, the domain using the proxy test will be determined to be equivalent provided that none of the proxy pretests have a difference that exceeds 0.25 SD.
- If all available proxy pretests have a difference between 0.05 and 0.25 SD and an adjustment for the pretest was not made by the authors in their analysis, the domain using the proxy test will be determined to be not equivalent.

The review requires that in a domain that requires statistical adjustments the adjustment is made only for that outcome. For example, if A, B, and C are available as pre- and post-intervention measures, and the pre-intervention difference in B requires statistical adjustment, only the analysis of outcome B must adjust for B.

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

- Percentage of students with low socioeconomic status
- Percentage of identified special education students
- Race or ethnicity
- Gender
- Percentage of students who speak English as a second language
- Teacher training or experience

Review team leadership may decide the differences indicate that the comparison group is not adequate for the purposes of this review.

Studies with an analytical model that includes the pretest score as a statistical covariate (e.g., ANCOVA) but do not demonstrate baseline equivalence will be considered when determining the evidence level, as there is evidence that these analyses can adequately control for selection (Fortson et al., 2012).

Outcomes

The *WWC Procedures and Standards Handbook* discusses the types of outcomes, criteria the outcome must meet, and how outcomes are reported by the WWC in Section 3 Subsection B 4 Outcome Eligibility and Reliability (pp. 16–19). This review follows the general guidance regarding reliability, outcomes measured at different points in time, impacts measured at different points in time, composite and subscale scores, subgroup findings, categorical ordinal measures, and estimated effects using imputed data.

Measures collected after the intervention ends are acceptable for this guide. To consistently examine effects across different interventions, measures administered closest to the end of the intervention will affect the level of evidence, but other outcome findings will also be reported in the guide appendix. Statistical adjustments to control for multiple comparisons will be computed within individual follow-up periods. Separate adjustments will be computed for the following follow-up periods, where appropriate: 2 weeks to 1 month, more than 1 to 3 months, more than 3 to 6 months, and more than 6 months. All outcomes within 2 weeks of the end of the intervention will be included in the immediate posttest adjustment. We will also report transfer outcomes—those that require students to apply concepts to new contexts—separately.

Study authors may use informal and experimenter-designed measures. Any experimenter-designed measure of reading ability is acceptable as long as the measure assesses students' reading skills and the measure is not overaligned with the treatment. To eliminate a measure from consideration, overalignment must be unambiguous; for example, an intervention where students write and edit a response to a particular prompt and then are assessed using the same prompt. When the review team is unable to determine the appropriateness of a measure, the panel chair will assist with determining whether the measure is acceptable.

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.

Citations

Fortson, Kenneth, Natalya Verbitsky-Savitz, Emma Kopa, and Philip Gleason. "Using an Experimental Evaluation of Charter Schools to Test Whether Nonexperimental Comparison Group Methods Can Replicate Experimental Impact Estimates." NCEE Technical Methods Report 2012-4019. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2012.