WWC EVIDENCE REVIEW PROTOCOL FOR BEGINNING READING INTERVENTIONS, VERSION 2.1[1]

Topic Area Focus

The What Works Clearinghouse (WWC) review for beginning readers focuses on reading interventions for students in grades K–3 (or ages 5–8) that are intended to increase skills in **alphabetics** (phonemic awareness, phonological awareness, letter identification, print awareness, and phonics), **reading fluency, comprehension** (vocabulary development and reading comprehension), or **general reading achievement** (see definitions below). Systematic reviews of evidence in this topic area address the following questions:

- Among interventions intended to provide basic literacy instruction, which ones improve reading skills (including alphabetics, reading fluency, comprehension, or general reading achievement) among students in grades K–3 (or ages 5–8)?
- Are some interventions more effective than others for certain types of reading skills?
- Are some interventions more effective for certain types of students, particularly students who have historically lagged behind in reading achievement?

Key Definitions

Alphabetics domain

***Phonemic awareness*.** 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 children learn how to read and spell by allowing them to combine or blend the separate sounds of a word to say the word (e.g., "/c/ /a/ /t/ – cat").

***Phonological awareness*.** 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*.** Letter identification refers to knowledge of the names of the letters of the alphabet.

[1] This protocol is aligned with the *WWC Procedures and Standards Handbook* (version 2.1).

***Print awareness*.** Print awareness refers to knowledge of concepts about print, such as (a) print carries a message; (b) print has conventions, such as directionality (left to right, top to bottom), differences between letters and words, distinctions between uppercase and lowercase, and punctuation; and (c) books have some common characteristics (e.g., author, title, front/back).

***Phonics*.** Phonics[2] refers to the (a) 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) ability to associate letters and letter combinations with sounds and blend them into syllables and words, and (c) understanding that this information can be used to decode or read words. Spelling is included in the review as an acceptable phonics outcome.

Reading Fluency domain

***Reading fluency*.** Fluency is the ability to read text accurately, automatically, and with expression (including appropriate pausing, response to punctuation, etc.) while extracting meaning from it.

Comprehension domain

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

***Reading comprehension*.** 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.

General Reading Achievement domain

***General reading achievement*.** 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) or provide some other type of summary score across domains, such as a "total reading score" on a standardized reading test, or grades in reading or language arts class.

[2] *Phonics* also refers to an instructional approach that focuses on the correspondence between sounds and symbols and often is used in contrast to whole-language instructional approaches. For the purposes of the Beginning Reading evidence reports, we use the term *phonics* as defined above, not as an instructional approach.

ELIGIBILITY CRITERIA AND EVIDENCE STANDARDS

Populations to be Included

The Beginning Reading topic area will review studies of interventions administered to children in grades K–3 (or ages 5–8).[3]

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 (e.g., a combined sample of students from grades K–4). The Adolescent Literacy topic area will review any studies that include fifth-grade students or higher (e.g., a combined sample of students from grades 3–6).

When study authors include a longitudinal sample of students who receive the intervention in grades covered by both the Beginning Reading and Adolescent Literacy topic areas, any studies in which the students receive the intervention in grade 4 (and lower) will be reviewed by the Beginning Reading topic area (e.g., a study in which receipt of the intervention spans 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 (e.g., a study in which receipt of the intervention spans grades 2–5).

Studies of children who are typically developing and those at risk for reading difficulties, including children with learning disabilities and low-income minority children, are considered for Beginning Reading reviews. To be included in our reviews, at least 50% of the students in each study must be general education students (which excludes students classified as English language learners [ELL] or limited English proficient [LEP] or those receiving special education services; these students are included in other WWC topic area reviews). For any studies that present analyses of subgroups of the study sample, the subgroup analyses also will be eligible for review in the Beginning Reading topic area.

These subgroups include the following:[4]

- Students of differing achievement levels (e.g., poor readers, underachievers)
- Students of different ages or grades (e.g., different levels of cognitive development)

[3] Studies that include students who are older than age 8 are included as long as they are in grades K–3. Interventions in pre-K also may be included if the children are age 5 or older. If authors do not provide the grade level of study students, we will use the age range of 5–8 to determine if the study is eligible for review within the Beginning Reading topic area.

[4] Subgroup analyses of students classified as needing special education services will be included in the other WWC topic areas focusing on special education students. Subgroup analyses of English language learners (ELL) or students classified as limited English proficient (LEP) will be included in the WWC English Language Learner topic area.

- Students from different socioeconomic backgrounds
- Students who are members of ethnic or racial minorities

Students living in high poverty and students from minority groups are of particular interest because these populations lag behind the population as a whole in reading achievement.

Types of Interventions to be Included

The specific interventions considered for inclusion are determined after an exhaustive search of the published and unpublished literature by the Beginning Reading team, as well as a review of nominations submitted to the WWC. Only research on interventions that are replicable (i.e., the intervention can be reproduced) are reviewed. The types of interventions included are as follows:

- Programs/products, such as
  - Comprehensive, non-textbook-based programs—some are curriculum based, others focus on staff development—intended to serve as a school's primary K–3 literacy instruction program, such as *Exemplary Center for Reading Instruction*, *Direct Instruction* (using *SRA Reading Mastery* materials), *Success for All*®, *Voyager Universal Literacy Program*®, and *Waterford Early Reading Program*™;
  - Supplemental K–3 reading programs intended to enhance whole-school literacy;
  - K–3 reading basals/textbooks intended for whole-school, whole-classroom use, such as Houghton-Mifflin, SRA/McGraw-Hill Reading Mastery, SRA/McGraw-Hill Open Court, Harcourt; and
  - Software designed to improve reading skills.
- Practices (e.g., reading aloud, home literacy environments)
- Policies (e.g., a schoolwide literacy initiative)
- Variations across programs, products, practices, and policies that include
  - Targeting to specific populations (e.g., readers below grade level, at-risk students)
  - The intention to be a school's primary literacy instruction program versus a supplemental reading program
  - Relative emphasis on implementing a packaged curriculum versus provision of professional development
  - Relative emphasis on a phonics versus whole language versus a "balanced" approach

  - Relative emphasis on home–school connections
  - Different level of implementation (national, statewide, districtwide, schoolwide, individual grades, whole group, small group, one-on-one)
  - Different medium/media through which program is implemented (e.g., basal readers, computer software)
  - Relative emphasis on enhancing specific beginning reading outcomes (e.g., phonemic awareness and phonics skills versus text comprehension skills)

- "Branded" and "non-branded" interventions. Branded interventions are commercial 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)
    - Sells or distributes the intervention
  - Replicability: packaged or otherwise available for distribution/use beyond a single site
  - Trademark

Elements of Intervention Replicability

The following characteristics of an intervention must be documented to reliably replicate the intervention with different participants, in other settings, and at other times:

- Intervention description
  - Skill(s) being targeted
  - Approach to enhancing the skill(s)
  - Targeted population
  - Unit of delivery of the intervention (i.e., whole group, small group, or individual student)
  - Medium/media of delivery of the intervention (i.e., teacher-led instruction or software)

- Intervention duration: amount of time to implement the intervention
- Description of individuals delivering the intervention

An intervention's effectiveness may vary by location, and a study that claims to test the effectiveness of an intervention should attempt to examine the effects of the intervention across different settings, such as:

- Location types (e.g., urban, rural, suburban)
- School types (e.g., public, private, parochial)
- Classroom types (e.g., regular, inclusion)

Types of Research Studies to be Included

The study must have been publicly released in 1983 or later and obtained by the WWC prior to drafting the intervention report. We believe this time frame adequately represents the current status of the field and allows for a manageable project scope.

The Beginning Reading literature search focuses on studies involving reading programs, practices, policies, and products to improve the reading skills for children in grades K–3. To be included in the review, a study must meet several relevancy criteria:

- ***Topic relevance.*** The study has to be about reading or literacy, focusing on alphabetics, reading fluency, comprehension, or general reading achievement. The study is required to focus on the effects of an intervention, not on
  - individual differences (e.g., correlational studies examining the covariance between reading speed and performance on a reading test; studies focusing on brain functions or structures)
  - assessment (e.g., on properties of an instrument)
- ***Time frame relevance.*** The study has to have been published in 1983 or later. Because of the difficulty in reaching authors of conference papers, we altered our time frame for those searches from the first version of the Beginning Reading review protocol (version 1.0) and limited our review of conference proceedings to those that took place from 1998 to 2005.
- ***Language relevance.*** The study should be written in English.
  - Studies that are written in English but focus on literacy in other languages are excluded from the review.

- ***Sample relevance.*** The sample must include students in grades K, 1, 2, or 3 learning to read in English.
  - The intervention must have taken place in grades K, 1, 2, or 3; outcomes may be measured in grades K–3 or later.
  - Studies that focus exclusively or primarily on early literacy among English language learners are excluded from consideration; these students are included in other WWC topic area reviews.
  - Studies that focus exclusively or primarily on students classified as needing special education services or on ELL or LEP students are excluded from consideration.[5]
- ***Study design relevance.*** Study designs that are eligible for review are limited to
  - Empirical studies using quantitative methods and inferential statistical analysis; and
  - Studies that take the form of a randomized controlled trial (RCT), a regression-discontinuity design (RDD), a quasi-experimental design (QED), or a single-case design (SCD).
- ***Outcome relevance.*** The study is required to
  - Focus on student outcomes (not teacher or other outcomes);
  - Focus on students' literacy-related outcomes (not non-literacy-related outcomes, such as motivation or engagement); and
  - Include at least one relevant outcome[6] that demonstrates adequate face validity or reliability.

Outcomes Included and Reliability Standards

Alphabetics (phonemic awareness, phonological awareness, letter identification, print knowledge, and phonics), reading fluency, comprehension (vocabulary development or reading comprehension), and general reading achievement are our primary outcomes of interest. Although we recognize the importance of motivation and attitudes toward reading, we have focused this review solely on student achievement outcomes.

[5] Studies in which at least 50% of the sample is classified as needing special education services will be reviewed in other WWC topic areas focusing on special education students. Studies in which at least 50% of the sample is classified as ELL or LEP will be reviewed in the ELL topic area.

[6] A relevant outcome is an outcome that falls in one of the domains listed on pages 1 and 2 of this document (in the "Key Definitions" section).

These literacy skills may be measured by standardized achievement tests, by researcher- or teacher-developed assessments, and by post-intervention class grades.

Reliability of outcome measures (internal consistency, temporal stability/test-retest reliability, and inter-rater reliability) will be assessed using the following standards determined by the WWC Standards and the Statistics, Technical, and Analysis Team (STAT):

For group design studies (RCT, RDD, or QED):

- Internal consistency score reliability: minimum of 0.60
- Temporal stability/test-retest score reliability: minimum of 0.40
- Inter-rater score reliability: minimum of 0.50 (percentage agreement, correlation, Kappa)

For single-case design research (SCD):

- The outcome variable must be measured systematically over time by more than one assessor, and the study needs to collect inter-assessor agreement in all phases and at least 20% of all sessions (total across phases) for each condition (e.g., baseline, intervention).

According to WWC standards, SCD studies must demonstrate reliability of outcomes through an inter-observer assessment (IOA; http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19, Appendix F). Inter-observer reliability is particularly relevant for measures that require an observer to score another person's behavior or complete a rating or checklist describing the behavior observed. Reading outcomes that involve written responses need not meet the SCD IOA requirement if the principal investigator (PI) determines that the responses can be scored by a single coder with a high degree of reliability. An example of outcomes that would not require IOA is written responses to a reading test. An example that would require IOA is coding of a written paragraph intended to demonstrate the quality of the student's narrative writing.

In addition to the reliability with which an outcome is measured, the PI will determine whether an outcome measure is too closely aligned with an intervention to be considered for inclusion in the review. An outcome may be determined to be too closely aligned with an intervention if the group that received the intervention was directly exposed to the outcome, such as a particular passage of text or vocabulary words, and the comparison group was not.

Reporting Outcomes in Intervention Reports

A reading intervention may have an immediate effect as well as a longer-term effect on student achievement. Thus, outcomes measured at the end of an intervention (posttest), as well as those measured any time thereafter (follow-ups), are relevant outcomes for review**.** As a rule, the end-of-intervention (posttest) outcome measure will be used to determine the WWC effectiveness rating and, therefore, will be reported in Appendix C of the intervention report. The follow-up outcome measures will generally be reported in Appendix D.

An effect can also be measured after different amounts of time of a reading intervention's implementation (for example, after 1, 2, or 3 years of intervention implementation). As a rule, the outcome measure that reflects the maximum exposure to the intervention (for example, after three years of the intervention's implementation) will be used to determine the WWC effectiveness rating (and, therefore, will be reported in Appendix C of the intervention report). The intermediate outcome measures (for example, after 1 and 2 years of implementation) will generally be reported in Appendix D. However, the PI may decide that intermediate outcome measures should contribute to the WWC effectiveness rating. The PI will consider factors such as the prevalence of findings across implementation levels and the design of the intervention when making this determination.

A study author can also report *both* a total test score *and* subtest scores (for example, the total CAT/5 reading score *and* the Vocabulary and Comprehension subtests). As a rule, the total score will be used to determine the WWC effectiveness rating (and, therefore, will be reported in Appendix C of the intervention report). The subtest scores will generally be reported in Appendix D. However, if the total score is not statistically significant and subtest findings are significant (or some of them are), then we will report the *subtest* scores in Appendix C and the *total* scores in Appendix D, so the subtest scores will factor into the WWC effectiveness rating.

For studies that also present analyses of subgroups of the study sample, the entire study sample findings will be used to determine the WWC effectiveness rating (and, therefore, will be reported in Appendix C of the intervention report). The subgroup analyses of interest for the Beginning Reading topic area will generally be reported in Appendix D.

Design Ratings

Sample attrition is a key factor in determining the WWC rating for RCTs. Baseline equivalence on measures of the outcome variable or factors correlated with the outcome measure is a key factor in determining the WWC rating for QEDs and RCTs with high attrition.

Attrition in RCTs. The WWC considers both the overall sample attrition rate and the differential in sample attrition between the intervention and comparison groups, as both contribute to the potential bias of the estimated effect of an intervention. The WWC has established conservative and liberal standards for acceptable levels of attrition. The conservative standards are applied in cases where the PI has reason to believe that much of the attrition is endogenous to the intervention reviewed—for example, high school students choosing whether or not to participate in a drop-out prevention program. The liberal standards are applied in cases where the PI has reason to believe that much of the attrition is exogenous to the intervention reviewed (e.g., in cases where movement of young children in and out of school districts is due to family mobility). Attrition rates are based on the number of sample cases used in the analysis sample with measured, as opposed to imputed, values of the outcome measures.

The Beginning Reading review uses the liberal standard. Table 1 presents the maximum difference in the attrition rates for the intervention and comparison groups that is acceptable for a given level of overall sample attrition. The empirical basis for these thresholds is described in Appendix A of the *WWC Procedures and Standards Handbook,* version 2.1.

Studies based on cluster random assignment designs must meet attrition standards for both the study sample units that were assigned to intervention or comparison group status (e.g., schools or districts) and the study sample units for analysis (e.g., typically, students). In applying the attrition standards to the subcluster level (e.g., students), the denominator for the attrition calculation includes only sample members in the clusters that remained in the study sample.

RCTs with combinations of overall and differential attrition rates that exceed the applicable threshold, based on the applicable standard, must demonstrate baseline equivalence of the analysis sample, or, if nonequivalence falls within the allowable range, statistically control for the nonequivalence, in order to receive the second-highest rating: *meets WWC evidence standards with reservations*. See the Baseline Equivalence section for more details.

Table 1: Attrition Standards for Randomized Controlled Trials

Highest Level of Differential Attrition Allowable to Meet the Attrition Standard Under the Liberal Attrition Standard

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

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*. Baseline equivalence is examined on measures of the outcomes or baseline measures that are expected to be highly correlated with these outcomes. For the Beginning Reading review, these variables are a pretest of an acceptable outcome measure, which can come from any of the four outcome domains:

1. alphabetics (phonemic awareness, phonological awareness, letter identification, print knowledge, and phonics)

2. reading fluency

3. comprehension (vocabulary development and reading comprehension)

4. general reading achievement

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 measures of reading skills (in *any* of the four domains), the study does not meet the baseline equivalence standard, regardless of whether or not the impacts are estimated using models that control for baseline characteristics. Therefore, the study is rated *does not meet WWC evidence standards*.

Finally, when there is evidence that the populations being compared are drawn from very different settings (such as rural versus urban, or high-SES versus low-SES), these settings may be deemed too dissimilar to provide an adequate comparison. In these cases, the study is rated *does not meet WWC evidence standards*.[7]

[7] The Beginning Reading review team also will examine other baseline characteristics (when available) to assess baseline equivalence of studies. These characteristics include, but are not limited to: gender, IQ scores, race/ethnicity, percentage of English-as-a-second-language students, measures of underserved population status, special education, school location, and average class size. The provision of all such information, however, is not a requirement of the review.

Statistical and Analytical Issues

Statistical properties of the data that are important in obtaining an accurate estimate of an effect size include the following:

- For most statistics (including d-indexes), normal distribution and homogeneous variances are important properties.

- Odds-ratios have no required desirable properties except the minimum of five observations per cell.

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 that do not include statistical controls for the pretest in the analysis of effects of the intervention, the effect size estimates will be adjusted for differences in pre-intervention characteristics at baseline (if available) using a difference-in-differences method (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, although 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 was a mismatch between the unit of assignment and the unit of analysis, and (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 (i.e., 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 Beginning Reading review is 0.20. For an explanation of the clustering correction, see Appendix C of the *Handbook*. All standards apply to overall findings and analyses of subsamples of students.

LITERATURE SEARCH METHODOLOGY

Collecting and Screening Studies

The WWC literature search is comprehensive and systematic. Detailed protocols guide the entire literature search process. At the beginning of the process, relevant journals, organizations, and experts are identified. The WWC searches core sources and additional topic-specific sources identified by the PI and the senior content advisor. The process is fully and publicly documented. For the Beginning Reading topic area, different search strategies were used for the "original" searches conducted from 2002 to 2007 and "new" searches conducted in 2007 and 2008.

Searches Conducted from 2002 to 2007

From 2002 to 2007, trained WWC staff members used the following strategies in collecting studies.

Databases

This is the core list of electronic databases that were searched across topics:

- ***ERIC.*** Funded by the U.S. Department of Education (ED), ERIC is a nationwide information network that acquires, catalogs, summarizes, and provides access to education information from all sources. All ED publications are included in its inventory.

- ***PsycINFO.*** PsycINFO contains more than 1.8 million citations and summaries of journal articles, book chapters, books, dissertations, and technical reports, all in the field of psychology. Journal coverage, which dates back to the 1800s, includes international material selected from more than 1,700 periodicals in more than 30 languages. More than 60,000 records are added each year.

- ***Campbell Collaboration.*** C2-SPECTR (Social, Psychological, Educational, and Criminological Trials Register) is a registry of more than 10,000 randomized and possibly randomized trials in education, social work and welfare, and criminal justice.

- ***Dissertation Abstracts.*** As described by Dialog, Dissertation Abstracts is a definitive subject, title, and author guide to virtually every American dissertation accepted at an accredited institution since 1861. Selected master's theses have been included since 1962. In addition, since 1988, the database includes citations for dissertations from 50 British universities that have been collected by and filmed at the British Library Document Supply Centre. Beginning with Dissertation Abstracts International, Volume 49, Number 2 (spring 1988), citations and abstracts from Section C (DAIC), Worldwide Dissertations (formerly European Dissertations), have been included in the file. Abstracts are included for doctoral records from July 1980 (Dissertation Abstracts International, Volume 41, Number 1) to the present. Abstracts are included for master's theses from spring 1988 (Masters Abstracts, Volume 26, Number 1) to the present.

- ***Sociological Collection.*** This database provides coverage of more than 500 full-text journals, including nearly 500 peer-reviewed titles. Sociological Collection offers information in all areas of sociology, including social behavior, human tendencies, interaction, relationships, community development, culture, and social structure. This database is updated daily via EBSCO host.

- ***Professional Development Collection.*** Designed for professional educators, this database provides a highly specialized collection of more than 500 full-text journals, including nearly 350 peer-reviewed titles. Professional Development Collection is the most comprehensive collection of full-text education journals in the world.

- ***Wilson Education Abstracts PlusText.*** Wilson Education Abstracts PlusText, also known as Education PlusText, combines abstracts and indexing from H. W. Wilson's Education Abstracts database with thousands of full-text and full-image articles. The database includes indexing and abstracts for articles published by more than 400 journals cited in H. W. Wilson's Education Abstracts database. It also includes full-text and full-image coverage for more than 175 of the sources. Its overall dates of coverage are 1994 to the present. Special education, adult education, home schooling, and language and linguistics are just a few of the hundreds of topics users can research in the database.

Search Parameters

After the identification of the topics for review, the project coordinator and the librarians initiated the search using keywords and search terms for each database. The senior content advisor reviewed and supplemented the list with additional keywords and search terms. Table 2 displays the list of keywords used for these Beginning Reading electronic searches.

Table 2. Beginning Reading Keywords Used for Electronic Searches

Keywords	ERIC Thesaurus Term(s)	PsycInfo Thesaurus Term(s)	Sociological Collection	Professional Development Collection	Dissertation Abstracts
Literacy	(R) Basic skills, alphabet (letters), literacy education, reading skills	(R) Reading skills, reading education, reading development, literacy programs	(OT) Information literacy	(OT) Information literacy	Use keywords from Keywords column as needed. There is an Education, reading subject category (Descriptor code: 0535)
Reading skills	(B) Language skills, reading ability	(N) Reading comprehension, reading speed (R) Sight vocabulary, word recognition	(SU) Reading comprehension, reading skills competency tests, ability testing	(SU) Reading comprehension, reading skills competency tests, ability testing	Use keywords from Keywords column as needed
Literacy instruction	(UT) Literacy education, (N) Basal reading, remedial reading (R) Reading instruction	Literacy programs (R) Reading skills, reading education, literacy	(SU) Literacy instruction	(SU) Literacy instruction	Use keywords from Keyword column as needed

Keywords	ERIC Thesaurus Term(s)	PsycInfo Thesaurus Term(s)	Sociological Collection	Professional Development Collection	Dissertation Abstracts
K–3	Kindergarten (B) Instructional program divisions, (NT) Kindergarten, grade 1, grade 2, grade 3	Primary school students	(SU) Grades K–3, kindergarten, first grade (education), second grade (education), third grade (education)	(SU) Grades K–3, kindergarten, first grade (education), second grade (education), third grade (education)	Use keywords from Keywords column as needed
Reading comprehension	Reading comprehension, (BT) Reading skills, comprehension (R) Reading strategies, readability, reading rate	Reading comprehension, (BT) Reading skills, verbal comprehension (R) Readability, reading	(SU) Reading comprehension	(SU) Reading comprehension	Use keywords from Keywords column as needed
Vocabulary development	Vocabulary development, (R) Lexicography, verbal development (S) Vocabulary building	(UT) Vocalization, (R) Communication, oral communication, verbal communication	(SU) Vocabulary, reading	(SU) Vocabulary, reading	Use keywords from Keyword column as needed
Phonemics	Phonemics (BT) Phonology, (NT) phonemes	Found: Phonetics (B) Phonology (R) Articulation (speech), phonemes	(SU) Reading, spelling ability, phonemics	(SU) Reading, spelling ability, phonemics	Use keywords from Keywords column as needed
Phonics	Phonics, (BT) Phonetics, (R) Aural learning, word study skills	Phonics (R) Reading education	(SU) Phonics, reading	(SU) Phonics, reading	Use keywords from Keywords column as needed
Reading fluency (term not in our print copy of ERIC thesaurus– added online 6/20/02)	Reading fluency, (R) Reading comprehension	Found: Reading materials (R) Reading, readability	NOT FOUND (UT) readability	NOT FOUND (UT) readability	Use keywords from Keywords column as needed
Intervention	Intervention, (N) Early intervention (R) Disabilities, disadvantaged, educational therapy	No specific term	No specific term	No specific term	Use keywords from Keywords column as needed
Instructional effectiveness	(R) Instructional improvement, program effectiveness, administrator effectiveness, curriculum evaluation educational quality, outcomes of evaluation	Found: Instructional media	No specific term	No specific term	Use keywords from Keywords column as needed

Keywords	ERIC Thesaurus Term(s)	PsycInfo Thesaurus Term(s)	Sociological Collection	Professional Development Collection	Dissertation Abstracts
Elementary school children	(UT) Elementary school students	Found: Elementary school students (N) Primary school students	No specific term	No specific term	Use keywords from Keywords column as needed
Assignments	(B) Instruction, (R) Homework, reading assignments	No specific term	(SU) Assignments	(SU) Assignments	Use keywords from Keywords column as needed
Reading achievement	Reading achievement (R) Reading failure, reading improvement, reading skills, achievement gains	Reading achievement (B) Academic achievement (R) Reading	(SU) Reading, literacy	(SU) Reading, literacy	Use keywords from Keywords column as needed
Improvement	(N) Reading improvement, speech improvement (R) Improvement programs, success	No specific term	No specific term	No specific term	Use keywords from Keywords column as needed
Instructional strategies	(UT) Educational strategies	No specific term	(SU) Instructional strategies	(SU) Instructional strategies	Use keywords from Keywords column as needed
Educational strategies	Educational strategies (BT) Educational methods, (R) Instructional design, learning strategies	No specific term	(SU) Educational strategy	(SU) Educational strategy	Use keywords from Keywords column as needed
Instructional programs (print copy deems "invalid descriptor")	Use Programs?	No specific term	No specific term	No specific term	Use keywords from Keywords column as needed
Instructional materials	(N) Courseware, learning modules, textbooks, workbooks, protocol materials (R) Reading materials, educational games, educational resources, instructional effectiveness, material development	Found: instructional media (N) Reading materials, textbooks	(SU) Educational tests and measurements, instructional materials industry	(SU) Educational tests and measurements, instructional materials industry	Use keywords from Keywords column as needed
Schoolwork	(UT) Assignments	No specific term	(SU) Education, schools	(SU) Education, schools	Use keywords from Keywords column as needed

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. Table 3 displays the list of 29 journals used for the Beginning Reading hand searches from 2002 to 2007. Six of these journals are topic specific and emphasize reading outcomes.

Table 3. List of Hand-Searched Journals[8]

American Educational Research Journal	Journal of Research and Development in Education
American Journal of Education	Journal of Special Education*
American Journal of Evaluation	Learning Disabilities: Research and Practice*
Educational Evaluation and Policy Analysis	Learning Disability Quarterly*
Educational Research and Evaluation	Peabody Journal of Education
Educational Researcher	Phi Delta Kappa
Effective School Practices	Reading Research Quarterly*
Evaluation and Program Planning	Review of Research in Education
Evaluation Review	School Effectiveness and School Improvement
Harvard Educational Review	Scientific Studies of Reading*
Journal of Education	Social Psychology of Education
Journal of Educational Psychology	Sociology of Education
Journal of Educational Research	Teachers College Record
Journal of Experimental Education	Urban Education
Journal of Literacy Research (aka Journal of Reading Behavior)*	

"Fugitive" or "Gray" Literature

Our search for fugitive or gray literature (difficult-to-find studies) encompassed seven strategies:

1. Review of the 1999–2003 conference programs for the annual meetings of the following organizations:

    - American Educational Research Association
    - American Psychological Association
    - American Sociological Association

[8] The asterisk (*) denotes a topic-specific journal.

- National Reading Conference (NRC)
- Society for Research on Child Development
- Society for the Scientific Studies of Reading

2. Public submissions
    - Materials submitted via the WWC website
    - Materials mailed or emailed directly to WWC staff
    - Materials given directly to WWC staff
3. Solicitations made to key researchers by the Evidence Report team
4. Requests for research made to developers of early literacy-related programs and interventions
    - Emailed or called approximately 130 developers to request voluntary participation
    - Developers identified through SEDL and the Northwest Regional Educational Laboratory (NWREL) websites; Florida Center for Reading Research (FCRR)
5. Checking prior reviews and research syntheses (i.e., using the reference lists of prior reviews and research syntheses to make sure that key studies have not been omitted)
    - National Reading Panel (NRP) Report
    - Other reviews identified through searches
6. Requests for research made via the NRC and Society for the Scientific Study of Reading (SSSR) listservs
7. Searches of organizations' websites
    - American Institutes for Research (AIR), Consortium for Policy Research in Education (CPRE), Mathematica, RAND, Urban Institute
    - American Federation of Teachers (AFT), National Education Association (NEA)
    - National Science Foundation (NSF), National Institute of Child Health & Human Development (NICHD)
    - Federally-funded research centers and programs (National Institute for Literacy [NIFL], Center for the Improvement of Early Reading Achievement [CIERA])

Searches Conducted Since 2007

New searches began in 2007 for studies published in 2005 or later. These searches used a different combination of databases and search parameters.

Databases

For the new searches, four of the databases were the same as those used in the original search: ERIC, PsycInfo, Dissertation Abstracts, and SocIndex (formerly called Sociological Collection). The three other databases from the original search (Campbell Collaboration, Professional Development Collection, and Wilson Education Abstracts) were replaced with two new databases:

- ***WorldCat.*** WorldCat is the world's largest network of library content and services. It allows users to simultaneously search the catalogs of more than 10,000 libraries, containing more than 1.2 billion books, dissertations, articles, CDs, and other media.

- ***Google Scholar.*** Google Scholar provides a simple way to broadly search for scholarly literature. From one place, users can search across many disciplines and sources: peer-reviewed papers, theses, books, abstracts, and articles from academic publishers, professional societies, preprint repositories, universities, and other scholarly organizations.

In addition, the websites of all the developers of literacy-related programs/products and interventions for any research or implementation reports were searched.

Search Parameters

The new searches differed from the original (2002–2007) ones by focusing on intervention names instead of general keywords. In particular, searches were conducted for (a) interventions with studies awaiting review at the time of expiration of the prior WWC contract (in July 2007), (b) interventions with studies that met WWC evidence standards with or without reservations under the prior WWC contract (2002–2007), and (c) additional interventions identified as high priority by the PI. All references resulting from these searches were screened for eligibility.