REVIEW PROTOCOL FOR THE READING COMPREHENSION PRACTICE GUIDE

(This guide is subject to WWC version 2.0 standards)

The Reading Comprehension for Beginning Readers practice guide is designed to provide concrete, actionable, research-based suggestions for classroom and school practices in grades K through 3 that can improve reading comprehension outcomes. Given the breadth of comprehension as a topic, the range of competing definitions in the field for comprehension, and the diverse needs of students, the panel will establish clear parameters for the scope of the guide, including the literature it considers when making research-based recommendations in order to produce a guide that is manageably-sized for practitioners.

Screening Criteria

Time Frame
Relevant to: Screening Criteria
Protocol: **1989–Present** (specific highly-related studies from 1978–1988 may also be reviewed upon request by the panel)

Age and/or Grade Range, and Ineligible Subgroups
Relevant to: Screening Criteria
Protocol:
Kindergarten through 3rd grade students
- Studies of students in grades K through 3 (ages 5 through 9), or in any subset of these grades are included.
- Studies that contain students in other grades are not included unless (1) study results disaggregate the results of students in eligible grades or (2) students in eligible grades represent the majority of the aggregated mixed-age sample.

Focus on English-speaking students without identified disabilities. The panel will not focus on, nor will it issue recommendations centered on, comprehension instruction and practices that target English language learners (ELLs) or students with learning disabilities. While comprehension instruction is important for those two groups of students, the needs of the two groups are distinct from most mainstream students and the panel will not focus on the two groups in order to limit the scope of the practice guide (though the panel will review and consider studies that include these populations). Further, the panel will restrict its recommendations to research-supported practices administered in English.

English-speaking students
- Studies of students that speak English and another language are included
- Studies with primarily students who have limited English proficiency are included (though reviewers should focus on results that disaggregate the English-speakers from other students if such disaggregated results are available, but reporting aggregated results is acceptable). In the case of studies with primarily (more than half) English language learners, reviewers should note the composition of the sample in the Additional Notes section of the study tracker.

Students without identified disabilities
Studies that focus solely on primarily students who have identified disabilities are not included. 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 are included. For studies that include both regular-track students and students with identified disabilities, reviewers should focus on results that disaggregate the students with disabilities from other students if such disaggregated results are available, but reporting aggregated results is acceptable.

Location and Language
Relevant to: Screening Criteria
Protocol:
Studies of a reading comprehension intervention that was conducted in English (with primarily English-speaking students—see above) are included. The study need not be set in the United States.

Study Design
Relevant to: Screening Criteria
Protocol: **RCTs, QEDs, and Regression Discontinuity Designs are eligible for review. Correlational and Qualitative studies will be held for future consideration.**

Regression Discontinuity Designs (RDDs) are eligible for review. Until WWC RDD standards are disseminated, please make note that the study is an RDD and the evidence coordinator will have it evaluated by an expert in RDDs.
Details on eligible interventions and comparisons:
- Intervention groups may receive "bundled" interventions (i.e., the intervention may be multi-faceted)
- Studies that compare multiple interventions are eligible for review (i.e., the comparison group does not have to be "business as usual")
- Studies comparing multiple levels of intervention are eligible (for example, Intervention A might be compared with Intervention A+B)

Completing SRGs with more than two groups: Studies may include three or more eligible groups (for example, Intervention A, Intervention B and a Comparison Group). If this is the case, reviewers will evaluate whether the study meets standards for each contrast (i.e., pair of groups). If any contrast meets standards, the study will be rated as *meets standards* (or *meet with reservations*). The SRG and Study Tracker Notes should clearly indicate any contrasts that do not meet standards, or that have a different rating from the main rating indicated in the study tracker. Where there are more than two groups, Reviewers will complete separate calculations on differential attrition for each contrast, and separate versions of the SRG Table 3 (effect size calculations) for each contrast (Note: multiple comparison adjustments will be made across all groups. See "Multiple Comparisons" below).

Correlational and Qualitative Research. Correlational and qualitative studies (particularly large or well designed ones) may be used in a confirmatory way to bolster evidence where there are

few studies or weakly designed studies. Studies with this design may be flagged for later summarization, but will not be assigned as part of the formal review process. The panel will determine whether studies with these designs should be considered after assessing the direction of emerging panel recommendations. While this type of research cannot move a claim to strong evidence, the panel will not rule them out without first considering the recommendations it wishes to put forward.

Attrition Guidelines for RCTs
Relevant to: Full Review
Reviews for this practice guide make optimistic assumptions about attrition (using the higher of the two WWC attrition thresholds) because, in the reading comprehension area, it is unlikely that overall or differential attrition will be due to intervention status (rather than to mobility among students). The SRG for this guide is programmed to assist reviewers with determining whether attrition in their study is too high in an RCT.

Baseline Characteristics for QED Equivalence (and RCTs with high attrition)
Relevant to: Full Review
QEDs must demonstrate baseline equivalence (0.25 standard deviation or less) of the analysis sample on pretest measures of the outcomes in order to meet standards with reservations. Failure to demonstrate baseline equivalence on <u>comprehension or another appropriate measure of baseline equivalence</u> will cause the study to fail to meet standards.

Studies that present baseline differences greater than 0.5 standard deviation on other important outcomes, such as demographics (specifically, free-lunch status, ELL status, and gender) or teacher training/experience may also fail to meet standards.

In order to meet standards with reservations, QEDs that demonstrate sufficient baseline equivalence of the analysis sample at pretest must also include appropriate statistical adjustments for baseline levels when comparing posttest outcomes.

Practice guide schedules are too limited to allow for author queries. Therefore, if, <u>demographics or teacher characteristics</u> are not mentioned in the study, reviewers will give the benefit of the doubt that those characteristics are equivalent or a non-issue (for example, a study with no ELL students may not discuss baseline equivalence on that characteristic, but should not be excluded from review because of it). **When giving QEDs the benefit of the doubt in these cases, reviewers must note in their SRGS and in the Study Tracker "Additional Notes" the specific characteristics that are assumed equivalent.**

Description of Intervention
Relevant to: Full Review
Protocol:
Definition of "Intervention." The guide will consider studies of branded comprehensive or supplemental curricula or effective and replicable strategies for teaching reading comprehension to students in Kindergarten through 3rd grade. These may include strategies or curricula used by

teachers in classrooms, those used by reading specialists in the school, or those for use by paraprofessional educators, tutors, or parents.

Relationship of Interventions to the Panel Recommendations. The proposed practice guide will examine practices to improve basic comprehension (i.e., extracting meaning from text) as well as age-appropriate practices aimed at teaching children to ask questions, draw conclusions, and otherwise read critically.

Recommendations in the guide will center on questions like:
- How can teachers implement effective text comprehension instruction strategies within their existing classroom curriculum?
- Which text comprehension instructional practices are most effective with which grade or reading levels?
- How should teachers balance their instructional time between teaching text comprehension and other activities?
- How can teachers engage students in discussions or interactions that improve their comprehension?
- How should a teacher balance his/her interactions with approaches that require greater student autonomy (cooperative grouping, Book Club, etc.)?
- How can teachers monitor their own implementation of recommended practices to improve outcomes for students?

Recommendations will NOT address the following: professional development, teacher preparation, and textbook design issues.

Criteria for a Description. A thorough description of the practices or curricula each intervention and/or comparison group received is required to assist the panelists with successfully incorporating the reviewed evidence into the guide. However, no study should fail to meet standards due to an incomplete description. Reviewer summaries of intervention descriptions (for inclusion in the study tracker) should include the following factors, if available:
- Types and number of activities
- Lesson duration and number of lessons
- Time elapsed from pretest to posttest
- Information on the fidelity of implementation
- Information on the texts used (Expository? Narrative? Mixed? At what reading level? What length?)
- Training of staff delivering the intervention, whether person delivering the intervention is regular classroom teacher, researcher, or some other person

Outcomes and Domains for Multiple Comparison
Relevant to: Screening Criteria, Full Review
Protocol:

Defining Text Comprehension. Text comprehension refers to the understanding of the meaning of a passage and the context in which the words occur. For the purposes of this guide, reviewers will restrict their consideration to studies that have measures of student achievement in text

comprehension as outcomes. Studies not containing an outcome that evaluates student text comprehension will be excluded from the review. **Studies containing a text comprehension outcome should be screened in (provided that they are otherwise within the scope of the guide) even if the interventions or practices under study are not aimed specifically at reading comprehension.** The Evidence Coordinator will provide reviewers with a preliminary list of measures in the text comprehension domain. The Panel Chair and panelists will decide about the relevance of other outcome measures as studies that include them are located.

Study authors may use informal and experimenter-designed measures. Any experimenter measure of text recall, understanding, or learning is acceptable as long as the test being recalled was not part of the training set (this would constitute overalignment with the intervention), and as long as the measure taps students' grasp of information in the text rather than memorization. The Panel Chair can assist reviewers with determining which experimenter-designed measures are acceptable for assessing text comprehension. For these types of experimenter-designed measures that are included on the basis of face validity, the study should not be marked down if the author does not provide validity statistics and reviewers will assume sufficient validity for these measures based on the above panel guidance.

Other Domains: Other reading outcomes may be included in studies that also examine text comprehension outcomes. The panel will be interested in the impacts in these domains, but **only for interventions which show an impact on text comprehension.** If studies include outcome measures in domains other than text comprehension, reviewers should report outcomes on those measures in the SRG, but the panel may decide not to discuss those other outcomes after evaluating the outcomes on text evidence measures. However, when making multiple comparison adjustments, reviewers should consider the following as separate domains:
- *Alphabetics*
- *Vocabulary*
- *Fluency/Oral Reading*
- *Aggregate measures (include multiple domains)*
- *Related measures (oral language, spelling, listening, rapid automated naming, writing)*

Missing Information
Relevant to: Full Review
Protocol: **No author queries.** Because of the tight timeline for practice guides, authors are not contacted as part of the review process.

Any missing information should be noted on the SRG and in the Study Tracker.

Vague descriptions of study design: Studies may describe using intervention and comparison groups, but may not specify whether units were randomly assigned. If the study does not specifically describe assignment as random, reviewers will treat the study as a QED.

Vague descriptions of group formation: Studies may describe randomization of intervention and comparison groups, but may not specify the number of units in each group. If this is the case,

practice guide reviewers should contact Shannon Monahan or Emily Sama Martin—these instances will be handled on a case-by-case basis.

Mismatch between unit of assignment and unit of analysis. In some studies, the unit of assignment may not match the unit of analysis. For example, the researchers may assign teachers to conditions, but may assess the outcomes on the students of those teachers. In addition to following WWC procedures for cluster adjustments in these mismatch cases, reviewers should take special care to note how analysis groups were formed when the units of assignment were randomly assigned (for example, teachers were randomly assigned to condition and five students from each classroom were analyzed). To preserve random assignment, participants must be assigned to condition based only on chance and each participant must have a nonzero probability of being assigned to a condition. A study randomly assigning teachers and administering the intervention to five students selected prior to randomization or by an empirical method (for example, those who performed lowest on a standardized assessment) is considered an RCT because the principles of random assignment are preserved. But, a study randomly assigning teachers and asking the teachers to subsequently recommend or select (in some subjective way) the students in their class to participate in the study is considered a QED and must demonstrate baseline equivalence of the analysis sample of students.

Overalignment of measures. If reviewers cannot determine whether outcome measures are overaligned with the intervention, the study will be rated *uncertain*. The SRG and Study Tracker notes field should clearly describe that this is the reason for uncertainty. The Panel Chair will assess the measures used in the study and provide guidance to the reviewers.

Study Ratings
Relevant to: Full Review
Meets Standards—Follow WWC version 2.0 standards.

Meets Standards with Reservations—Follow WWC version 2.0 standards.

Uncertain—Follow WWC version 2.0 standards. If missing information *prevents reviewers from determining how a study should be rated*, the study should be marked as *uncertain.* If other information is missing that would enhance the study description in the SRG, but would not affect the rating of the study, the *uncertain* label should not be applied.

Does Not Meet Standards—Follow WWC version 2.0 standards.

Does Not Pass Screens—Follow this protocol with respect to timeframe, age/grade range, location of study, intervention content or type, outcomes measured, and study design.

Study for Support, Not Review—Apply this rating to studies with a design that is not eligible for review, but that have been recommended by a panelist (for example, key literature reviews or meta-analyses in the field). These studies will be assigned to staff (not necessarily reviewers) who will summarize them and draw out key messages that are relevant to the guide. Some studies not passing the screen, but related to a key point in the guide and providing a low level of

evidence, may be assigned this rating by the Evidence Coordinator after the study is screened out of the review process.

Multiple Comparisons
Relevant to: Full Review
Protocol: Reviewers will follow WWC guidelines for multiple comparison adjustment, making adjustments for all outcomes within the text comprehension domain.

Studies with Multiple Groups
Studies may include more than two groups. If that is the case, when adjusting for multiple comparisons, reviewers should count the total number of outcome/pair combinations for a given intervention. For example, consider a study that has five outcomes in text comprehension, and three groups (Intervention 1, Intervention 2, Comparison), where all groups have data on all five outcomes. The total number of groups for a multiple comparison adjustment on a spreadsheet calculating the effect of Intervention 1 will be 10: I1 vs I2–five outcomes ***plus*** I1 vs Comparison–five outcomes. Also, see the study design section of this protocol for more information on completing the SRG for studies with multiple groups. (If your study contains multiple age groups, some of which are outside the protocol, please discuss the correct adjustment procedure with your reconciler.)

Judging Significance After Multiple Comparison Adjustments
The version 2.0 SRG does not automatically calculate significance for you after a multiple comparison adjustment, because this decision requires some user judgment. The comparison that you make is both across the rows AND down the columns and the instruction is: Identify the largest p-value rank such that the p-value is less than or equal to the critical p-value. All findings with p-values smaller than or equal to that cutoff are significant, and all above are not significant.

In the case below: Outcome 1 has the largest p-value that is SIMULTANEOUSLY less than or equal to its' critical p-value. So outcome 1 becomes the "cut-off," and any p-value less than or equal to this cutoff (in this case Outcome 2) is also significant.

Outcome p-value   critical p-value
Outcome1
0.03

3
2
0.033
Outcome2
 0.02

3
1
0.017
Outcome3
 1.00

3
3
0.050
To keep track of your decision, it may be helpful to add information to column AV of the SRG noting whether the comparison is "sig" or "ns."

Collecting and Screening Studies

Sources for Studies—Searches conducted in 2002–2007

Databases

We will search the following electronic databases to locate studies for review:

1. **ERIC.** Funded by the U.S. Department of Education, 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.

2. **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 over 30 languages. More than 60,000 records are added each year.

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

4. **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 Masters 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 Document Supply Center. Beginning with DAIC Volume 49, Number 2 (Spring 1988), citations and abstracts from Section C, 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.

5. **Academic Search Premier.** This multi-disciplinary database provides full text for more than 4,500 journals, including full text for more than 3,700 peer-reviewed titles. PDF backfiles to 1975 or further are available for well over one hundred journals, and searchable cited references are provided for more than 1,000 titles.

6. **EconLit.** EconLit, the American Economic Association's electronic database, is the world's foremost source of references to economic literature. The database contains more than 785,000 records from 1969-present. EconLit covers virtually every area related to economics.

7. **Business Source Corporate.** Contains full text from nearly 3,000 quality business and economics magazines and journals (including full text of many only abstracted in other sources we search). Information in this database dates as far back as 1965.

8. **SocINDEX with Full Text.** SocINDEX with Full Text is the world's most comprehensive and highest quality sociology research database. The database features more than 1,986,000

records with subject headings from a 19,600+ term sociological thesaurus designed by subject experts and expert lexicographers. SocINDEX with Full Text contains full text for 708 journals dating back to 1908. This database also includes full text for more than 780 books and monographs, and full text for 9,333 conference papers.

9. ***EJS E-Journals.*** E-Journals from EBSCO host[®]: Find article-level access for thousands of E-Journals available through EBSCO's Electronic Journal Service (EJS). This resource covers journals MPR subscribes to.

10. ***Education Research Complete.*** Education Research Complete is the definitive online resource for education research. Topics covered include all levels of education from early childhood to higher education, and all educational specialties, such as multilingual education, health education, and testing. Education Research Complete provides indexing and abstracts for more than 1,840 journals, as well as full text for more than 950 journals, and includes full text for more than 81 books and monographs, and for numerous education-related conference papers.

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

13. ***Cochrane Central Register of Controlled Trials.*** Cochrane Controlled Trials Register is a bibliography of controlled trials identified by contributors to the Cochrane Collaboration and others, as part of an international effort to hand-search the world's journals and create an unbiased source of data for systematic reviews.

15. ***Database of Abstracts of Reviews of Effects.*** Database of Abstracts of Reviews of Effects (DARE) includes abstracts of published systematic reviews on the effects of health care from around the world, which have been critically analyzed according to a high standard of criteria. This database provides access to quality reviews in subjects for which a Cochrane review may not yet exist.

16. ***Cochrane Methodology Register.*** The *Cochrane Methodology Register* (CMR) is a bibliography of publications which report on methods used in the conduct of controlled trials. It includes journal articles, books and conference proceedings; these articles are taken from the *MEDLINE* database and from hand searches. The database contains studies of methods used in reviews and more general methodological studies which could be relevant to anyone preparing systematic reviews. CMR records contain the title of the article, information on where it was published (bibliographic details), and in some cases, a summary of the article. CMR is produced by the UK Cochrane Centre, on behalf of the Cochrane Methodology Review Group.

17. ***Google EdResearch.*** A custom Google search engine, searching ONLY the sites of organizations/teams requested by the searcher, and examining for relevance only 100 hits deep into the results of the searches in these databases. The current list of organizations included is:

Abt Associates	Institute for Public Policy and Social Research (IPPSR)
Alliance for Excellent Education	Johns Hopkins University School of Education
American Enterprise Institute	Learning Point Associates
American Institutes of Research	Linguistic Society of America (LSA)
American Speech-Language-Hearing Association (ASHA)	Mathematica Policy Research homepage
Appalachian Education Laboratory (Edvantia)	MDRC
Best Evidence Encyclopedia	Mid-continent Research for Education and Learning
Broad Foundation (Education)	National Association of State Boards of Education
Brookings Institution	National Autism Center - National Standards Project
Carnegie Corporation of New York	National Governors' Association
Center for Comprehensive School Reform and Improvement	National Reading Panel
Center for Data-Driven Reform in Education	Pacific Resources for Education and Learning (PREL)
Center for Research and Reform in Education	Public Education Network
Center for Research in Educational Policy (CREP)	Public Policy Research Institute at Texas A&M University
Center for Social Organization of Schools	Public/Private Ventures (PPV)
Center on Education Policy	RAND
Center on Instruction	Southwest Educational Development Laboratory (SEDL)
Chapin Hall Center for Children at the University of Chicago	SRI
Congressional Research Service (via OpenCRS.org)	Teachers of English to Speakers of Other Languages (TESOL)
Florida Center for Reading Research (FCCR)	Technical Assistance Center on Social Emotional Intervention for Young Children
Government Accountability Office (GAO)	The Education Resources Institute
Harvard Graduate School of Education	Thomas B. Fordham Institute
Heritage Foundation	U.S. Department of Education (includes Institute for Education Sciences, National Center for Special Education Research etc)
Hoover Institution	Urban Institute

Search Parameters. The Evidence Coordinator, Panel Assistance, and MPR library staff initiate a search using keywords and search terms for each database. The Panel Chair reviews and supplements the list with additional keywords and search terms. Table 1 displays the list of keywords used for the reading comprehension electronic searches.

Table 1. Reading Comprehension Keywords Used for Electronic Searches

Category	ID	Term
grade levels	S1	k-3 or kindergart* or (first and grade) or (second and grade) or (third and grade) or elementary
outcomes	S2	achievement* or improvement* or "instructional effectiveness" or outcome* or effect* or development or skill*
activity	S3	teach* or strateg* or instruction* or curricul* or approach* or monitor* or treatment or intervention*
combine terms	S4	(S1 and S2) or (S1 and S3)
topical terms	S5	reading or literacy
topical terms	S6	comprehension
combine terms	S7	4 and 5 and 6
topical terms	S8	reciprocal teaching AND S7
topical terms	S9	question* and (generat* or answer*) and S7
topical terms	S10	"text comprehension" and S7
topical terms	S11	(cognit* or metacognit* or "meta-cognit*" or "meta cognit*") and S7
topical terms	S12	((summar* and text) or (summar* and passage)) and S7
topical terms	S13	("explicit instruction" or "direct instruction") and S7
topical terms	S14	scaffold* and S7
topical terms	S15	(story and structur*) and S7
topical terms	S16	("prior knowledge" or "prior experience" or "schema theory") and S7
topical terms	S17	(cooperative and learning) and S7
topical terms	S18	"book club" or discussion or "cooperative grouping*" or listening
topical terms	S19	S18 and S7