# ELEMENTARY WRITING PRACTICE GUIDE
# REVIEW PROTOCOL VERSION 2.1

This protocol guided the review of research that informed the recommendations contained in the What Works Clearinghouse (WWC) practice guide "Teaching Elementary School Students to Be Effective Writers," published in June 2012.

This protocol was developed to reflect standard WWC procedures for reviewing research. The protocol also was informed by the objectives identified in the guide's formal purpose statement, which states the guide's goal is to provide evidence-based recommendations for improving elementary school students' writing. The purpose statement identifies key questions to be addressed by the panel, including:

1. How should teachers and schools create a climate to encourage writing?

2. How should teachers customize or differentiate writing instruction for students in different grades and of different writing abilities?

3. How should teachers help students acquire a process to think about what to write?

4. How much time should students spend on writing, and under what conditions?

5. How should teachers and schools employ technology to improve writing instruction?

The panel sought to identify research that examined these and other issues related to the quality of students' writing.

The review of research involved the following steps:

- The research staff searched the professional literature following the guidelines in this protocol to identify relevant studies. Additional studies were identified by the panel of experts.

- Studies were screened against eligibility criteria in this protocol to determine whether they were within the scope of the practice guide.

- Eligible studies were assessed against WWC evidence standards and the supplemental technical criteria specific to the Elementary Writing guide in this protocol.

  o Studies that met WWC evidence standards and were related to a recommendation within the guide were used to identify the strength of the evidence for each recommendation.

  o Studies that did not meet WWC evidence standards could be used to provide examples of practices. (Note: This differs from the procedures for WWC intervention reports, which report findings only for studies meeting WWC evidence standards.)

This protocol contains three parts: (a) procedures for conducting the literature search; (b) eligibility criteria for reviewing relevant studies; and (c) guidelines for handling technical issues

including attrition and group equivalence. Please refer to the *WWC Procedures and Standards Handbook (version 2.1)* for additional information.

## PROCEDURES FOR CONDUCTING THE LITERATURE SEARCH

The literature search for this practice guide involved a keyword search of multiple databases to identify studies examining the effectiveness of strategies for teaching writing to elementary school students.

### Keyword Search

*Primary Objective.* The primary objective was to identify relevant practices for review by (a) identifying practices with potentially eligible studies, and (b) determining the approximate number of eligible studies related to each practice.

*Search Strategy.* Keywords were selected that aimed to identify literature related to teaching writing to elementary school students. Keywords related to focus, setting, process, outcomes, grade levels, and study design were included to focus the search on literature that met the eligibility criteria for this review (see p. 5). The keyword list appears below. The list of databases that were searched appears in the next section.

- *Keywords that identified studies with a writing focus:* writing.

- *Keywords that identified interventions delivered in an educational setting:* school, education, student, child.

- *Keywords that identified process:* intervention, curriculum, program, strategy, instruction, teach, train, technique, approach, practice.

- *Keywords that identified studies with outcomes:* essay, instrument, portfolio assessment, sentence combining, fluency, voice, rubric, content learning, learning to write, writing to learn, word processing, paragraph, sentence, writing process, edit, proofread, error monitoring, error correction, peer revising, revising, composition, handwriting, spell, capitalize, punctuation, composing, mechanics, vocabulary, grammar, writing quality, analytic scoring, holistic scoring, length, number of words.

- *Keywords that identified relevant grade levels:* K–5, kindergarten, first grade, second grade, third grade, fourth grade, fifth grade, sixth grade, elementary.[1]

- *Keywords that identified studies with a relevant design:* control group, comparison group, matched groups, treatment, random, assignment, baseline, experiment, evaluation, impact, effectiveness, causal, posttest, post-test, pretest, pre-test, randomized controlled trial, RCT, quasi-experimental, quasi experimental, QED, regression discontinuity, changing criterion, intrasubject replication, multiple baseline, multi-element, multi

---

[1] Originally, the practice guide focused on students in grades K–5. The panel later decided to include students in grade 6 who were in an elementary setting, and studies of sixth graders were included in the review.

element, single case, single subject, ABAB, alternating treatment, simultaneous treatment, meta-analysis, meta analysis, reversal design, withdrawal design.

Multiple-word phrases listed above were searched as phrases. The keywords in each of the above categories were linked together with OR in a search so that they identified all articles that focused on any of the terms. The six sets of search terms were then linked together with AND in a search so that they identified all articles that focused on writing, were delivered in an educational setting, used an instructional process, had relevant outcomes, studied relevant grades, and used a relevant design. Finally, variations of words (e.g., "achieve*," to capture studies including the words "achieve" and "achievement") were searched to ensure that our search was as inclusive as possible.

## Databases

The search was conducted using the following databases:

- *Academic Search Premier.* This multidisciplinary 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 100 journals, and searchable cited references are provided for more than 1,000 titles.

- *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 to the present. EconLit covers virtually every area related to economics.

- *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.

- *EJS E-Journals.* E-journals from EBSCO host® provide article-level access to thousands of e-journals available through EBSCO's Electronic Journal Service (EJS).

- *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.

- ***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.

Results in Academic Search Premier, EconLit, Education Research Complete, ERIC, ESE E-Journals, and SocINDEX with Full Text were limited to scholarly (peer-reviewed) journals. Results in PsycINFO were limited to the "0100 Journal" publication type.

## "Fugitive" or "Grey" Literature

"Fugitive" or "grey" literature refers to studies that are not published commercially or are otherwise inaccessible through conventional literature searches. To be considered by the WWC, these studies must be available to the public. To identify "fugitive" or "grey" literature for this review, the review team solicited recommendations from panel members. Dissertations or master's theses were not included in the review unless specifically requested by a member of the panel.

## ELIGIBILITY CRITERIA AND EVIDENCE STANDARDS

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

## Populations to be Included

*Eligible grades*. Studies of elementary students in kindergarten through sixth grade (ages 5–12), or any subset of these grades, were included in the review. However, studies of sixth-grade students were only included if there was reason to believe the students were in an elementary school setting (K–8, K–6, or 4–6 schools were considered to be an elementary setting, but 6–8 or 6–12 schools were considered to be a middle school setting). If a study included sixth graders, but it was not clear from the study text whether the sixth graders were in an elementary or middle school setting, the review assumed the sixth graders were in an elementary setting. Studies of interventions that took place in the summer prior to entry into kindergarten were not included in the review; however, interventions that took place in the summer prior to entry into middle school were included.

Studies that contained students in other grades were not included unless (a) study results disaggregated the results of students in eligible grades, or (b) students in eligible grades represented the majority of the aggregated mixed-age sample. If the study did not make explicit the number of students in each grade, a study was included if 50% or more of the grades included in the sample fell within the eligible grade range.

*Special populations.* Though the guide did not focus on or issue recommendations that were specific to special populations, studies targeting or including non-English speaking students or

students identified for special education services were included in the review, and disaggregated results were reported when available .

- ***Studies of non-English speaking students.*** Studies of students that speak English and another language were included in the review. Studies primarily of students who had limited English proficiency were also included. Studies of students learning to write in languages other than English were included if the language was a first language and the educational setting was similar to educational settings in the United States (for example, studies of Spanish writing instruction delivered to Spanish-speaking students in Spain were included).

- ***Studies of students at risk for writing difficulties.*** Studies of students with identified disabilities or other characteristics that might make it more challenging for them to learn to write were included in the review. Other populations considered to be at risk for writing difficulties included students who find it difficult to regulate their behavior or students who struggle with skills related to writing, such as reading, spelling, or handwriting.

## Types of Practices and Interventions to be Included

The panel considered studies of branded curricula (including comprehensive or supplemental) or replicable strategies for teaching writing to students in elementary school. These included strategies or curricula used by teachers, writing specialists, paraprofessional educators, tutors, and parents.

## Types of Research Studies to be Included

To be included in the review, the study must have been written in English, and have met the following relevancy criteria:

***Topic relevance.*** The recommendations in the practice guide focused on instructional strategies that improve students' (a) basic writing skills, (b) ability to brainstorm and plan for a composition, and (c) ability to edit a composition through the use of corrective feedback and self-regulation strategies. Research studies had to be directly related to one of these topics.

***Time frame relevance.*** The study had to have been published between 1989 and June 2010; earlier or later work was reviewed if suggested by a panelist. The time frame was established in order to define a realistic scope of work for the review. The time frame helps focus the review on interventions that may be available to teachers today. It also helps ensure that effectiveness is characterized relative to conditions similar to those in schools today.

***Study design relevance.*** Randomized controlled trials (RCTs) and quasi-experimental designs (QEDs) were eligible for review. Regression discontinuity designs (RDDs) and single case designs (SCDs) were eligible for review using the WWC pilot standards for RDD and SCD research.

***Intervention and comparison group relevance.*** Eligible intervention and comparison groups included:

- Intervention groups that received "bundled" interventions (that is, the intervention may have been multi-faceted and included multiple components)

- Multiple levels of intervention (for example, Intervention A might have been compared with Intervention A+B)

- Multiple comparison groups, typically other interventions (the guide prioritized the comparison most relevant for a recommendation but may have used each of the comparisons or combined groups where appropriate)

**Types of Outcomes to be Included**

The panel was primarily concerned with interventions or practices designed to help children learn to write effectively (as opposed to those interventions where children write to learn more about a particular topic). The panel only considered measures of student ability based on original, student-written products (or "authentic writing"), because assessments that do not include authentic writing are not good predictors of writing quality. The panel made one exception to this rule: norm-referenced standardized tests of writing achievement. This exception was made because teachers are increasingly called upon to demonstrate improvement on these tests and are likely to be interested in interventions that have demonstrated impacts on these types of assessments.

The panel prioritized measures of **overall writing quality**, which measured the overall effectiveness of a student's writing. These measures may have taken into account assessments of intermediary outcome categories—including ideation, genre (or text) elements, mechanics, organization, output, sentence structure, vocabulary, and voice—in a single assessment of the quality of a piece of writing. Overall writing quality may have been assessed either analytically or holistically. Analytic writing quality referred to measures using scales for which multiple attributes of writing (e.g., mechanics, vocabulary, sentence structure, organization, ideation, and voice) were each judged separately and then summed to obtain a single score. Holistic writing quality referred to measures where the assessor made a single judgment about overall quality, considering a variety of attributes at the same time.

The panel also considered evidence of effectiveness in any of the outcome domains defined below:

- *Genre* elements, sometimes referred to as "text elements," assess the presence or quality of specific features typical of a particular genre. For example, the elements of a story might have included place, a starting event, action, and ending.

- *Ideation* referred to the development and quality of ideas students included in their writing. Qualitative measures of ideation included the overall richness and number of ideas in a composition. Quantitative measures included the number of different ideas.

- *Organization* assessed the structure of a composition. This could have included the connection between ideas in the text, as well as how well individual ideas were organized or connected to meet a writer's purpose (often referred to as "cohesiveness").

- ***Sentence structure*** assessed sentence correctness or sentence complexity. For example, a sentence-structure measurement might have counted the number of sentences in a composition that were syntactically correct.

- ***Vocabulary*** referred to the types of words used by the student in his or her writing. Vocabulary may have been assessed by counting specific types of words (e.g., the number of different words or the inclusion of content-specific words), or by examining the complexity of words (e.g., number of syllables).

- ***Voice,*** often referred to as "tone," "mood," or "style," tells the reader about the writer's personality in the composition. Voice typically was assessed by rating how well the student established mood, tone, style, or his or her individual personality in writing.

- ***Writing output*** referred to the actual quantity of text produced. Some examples of output measures included the number of sentences or the number of words in a composition.

Other information about or requirements for outcomes include the following:

***Experimenter-designed measures.*** Informal and experimenter-designed measures were eligible for review.

***Overalignment of outcomes.*** Outcome measures were determined to be overaligned with an intervention if the measure included similar materials to those used in the intervention or the measure was administered to the treatment group as part of the intervention. Outcome measures that were determined to be overaligned with an intervention were not eligible for review.

***Timing of outcome measurement.*** The panel was interested in whether improvements in writing quality could be sustained over time. Measures collected at the end of the intervention and anytime thereafter were acceptable for the purpose of this review. To be consistent in examining effects across different interventions, immediate post-intervention findings were prioritized, but delayed outcome findings were also included in the review. Statistical adjustments to control for multiple comparisons were computed within individual follow-up periods. Separate adjustments were 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 were included in the immediate posttest adjustment.

***Reliability.*** Given the subjective nature of many writing assessments, the reliability of outcome measures was assessed using inter-rater reliability (IRR) calculations. The panel set a minimum standard for IRR of a 0.6 correlation or 80% agreement within one point. When IRR was assessed as within two or more points, it was insufficient to meet the standards for this review. IRR must have been established on the study sample. In other words, "trained to reliability" was not sufficient for establishing IRR. IRR did not need to be established on the full study sample, however. If the authors did not specify the sample on which IRR was established, the review assumed it was assessed on the study sample. Studies were not required to document IRR on the study sample for approved standardized measures. IRR was not required for some measures that were more objective (for example, word count); however, these measures had to be approved on an individual basis by the panel chair.

# STATISTICAL AND TECHNICAL ISSUES

Eligible studies were assessed against WWC evidence standards, as described in the *WWC Procedures and Standards Handbook (version 2.1)*. Supplemental technical guidelines specific to the Elementary Writing guide are outlined in this section.

## Attrition in RCTs

The panel chair had reason to believe that much of the attrition in studies of elementary writing interventions is exogenous, thus optimistic assumptions about the relationship between attrition and the outcomes of interest were used for this guide. For more information, see Appendix A of the *WWC Handbook*.

## Group Equivalence in RCTs with High Attrition and QEDs

For this review, the studies should have demonstrated baseline equivalence in the domain of the outcome variable. However, some of the students in the eligible sample might have been unable to produce an original text at baseline. In cases where there were no eligible writing measures available at baseline, general cognitive ability, word reading ability, reading fluency, or reading comprehension measures were accepted to establish baseline equivalence of the intervention and comparison groups.

Equivalence was established on a domain-by-domain basis. Groups were considered equivalent in a particular domain if the reported differences in pre-intervention characteristics of the groups were less than or equal to one-quarter of the pooled standard deviation in the sample, regardless of statistical significance. However, if differences were greater than 0.05 standard deviations and less than or equal to one-quarter of the pooled standard deviation in the sample, the analysis must have controlled analytically for the individual-level pre-intervention characteristic(s) on which the groups differ (see Statistical and Analytical Issues below). If pre-intervention differences were greater than 0.25 for *any* of the listed characteristics in the same domain, the measures in that domain did not meet standards.

## Statistical and Analytical Issues

***Statistical controls.*** For RCTs, the effect size estimates were adjusted for differences in pre-intervention characteristics at baseline (if available) using a difference-in-differences method if the authors did not adjust for pretest (see Appendix B of the *WWC Handbook*).

***Adjustments to statistical significance.*** If a study included more than two groups relevant to the practice guide, when adjusting for multiple comparisons, reviewers counted the total number of outcome/pair combinations within a study relevant to a particular intervention. For example, in a study that has five outcomes in a domain, and three groups (Intervention, Comparison 1, Comparison 2) where all groups have data on all five outcomes, the total number of groups for a multiple comparison adjustment for Intervention will be ten (five outcomes in Intervention vs. Comparison 1 plus the five outcomes in Intervention vs. Comparison 2). See Appendix D of the *WWC Handbook* for the formulas the WWC uses to adjust for multiple comparisons.