

WWC EVIDENCE REVIEW PROTOCOL FOR HIGH SCHOOL MATHEMATICS INTERVENTIONS

VERSION 2.0 (5/2010)

Topic Area Focus

This What Works Clearinghouse (WWC) review focuses on mathematics interventions for high school students in grades 9–12 designed to impact student achievement, including curriculum-based interventions, instructional techniques, and products designed to deliver content and monitor student progress. Systematic reviews of evidence in this topic area address the following questions:

- Which interventions are effective in increasing the learning of mathematics content and skills among high school students?
- Are some interventions more effective than others for certain types of students, particularly students who are at risk of failure in mathematics?

Key Definitions

Mathematics Intervention. In this review, a mathematics intervention is defined as a replicable, materials-based instructional program that

- Is delivered to high school students
- Clearly delineates mathematics learning goals for students
- Is designed to directly affect student mathematics achievement

Mathematics Achievement Domain. Outcomes that fall in the mathematics achievement domain are those related to mathematics content and skills, commonly described as what students should know and be able to do. Mathematics content varies somewhat across curricula and grade levels, but generally includes pre-algebra, algebra, geometry, trigonometry, pre-calculus, and calculus. Mathematics skills involve the application of the learning of this content, as well as an understanding of mathematical concepts, procedures, and problem solving.

GENERAL INCLUSION CRITERIA

What Works Clearinghouse Criteria

Timeframe Relevance. The study must have been written between January 1989 and December 2008. This 20-year window is wide enough to allow for a baseline of data with regard to traditional curricula, National Assessment of Educational Progress (NAEP) trends, and standards-based curricula.

Study Design Relevance. The design must be an empirical study, using quantitative methods and inferential statistics, that includes a comparison group. Eligible designs include well-conducted randomized controlled trials (RCTs), quasi-experiments (QEDs) with matching or equating of student samples on a baseline student-level measure, regression discontinuity (RD) designs, and single-case (SC) designs.

High School Mathematics Criteria

Topic Area Relevance. The study must focus on the effects of a mathematics intervention on one or more measures of mathematics achievement.

Sample Relevance. The WWC High School Mathematics area reviews interventions for high school students from grade 9 through grade 12.

Geographic Relevance. The study must have been conducted in the United States (including the 50 states, the District of Columbia, territories, and tribal entities).

Outcome Relevance. The study must include at least one student achievement measure that demonstrates sufficient reliability or face validity.

SPECIFIC TOPIC PARAMETERS

Types of Populations to be Included

Defining Characteristics of the Target Population. The sample must include high school students, for whom “high school” is primarily defined as a school with any of the four grades from grade 9 through grade 12 and not explicitly classified as middle school. Students in grades lower than 9th grade are included in the review only if such students were classified in the study as high school students or were included along with students in any of the grades 9 through 12 in the study analysis sample; otherwise, students in grades prior to 9th grade fall within the scope of the topic area reviews for Middle School Mathematics. If general education students comprise less than 50% of the sample, the study will be covered in the appropriate special education topic area review.

Effectiveness of the Intervention Across Different Groups. An intervention’s effectiveness could vary by subgroups of student characteristics. Whether or not a study examines effects on subgroups does not affect its inclusion for review or the rating given to it. However, we will present in an appendix findings for subgroups of interest, provided that the subgroups are equivalent with regard to pretest and grade level. Student characteristics of interest for this review include the following:

- Baseline mathematics achievement
- Grade
- Gender
- Socioeconomic status (SES)
- Racial/ethnic breakdown
- Percentage of English as a second language students
- Percentage of bicultural students
- “At-risk” status (as provided by study authors)

Effectiveness of the Intervention Across Different Settings. An intervention’s effectiveness could vary by subgroups of setting characteristics. Whether or not a study examines effects on subgroups does not affect its inclusion for review or the rating given to it. However, we will present in an appendix findings for subgroups of interest, provided that the subgroups are equivalent with regard to pretest and grade level. School setting characteristics of interest for this review include the following:

- Location of the schools involved
- Homogeneous groupings of students
- School type (public, private, religious)
- School SES (e.g., Title I school)
- Average class size (small, medium, large)
- Average teacher characteristics (e.g., teacher education and experience)
- School size

Types of Interventions to be Included

Characteristics of High School Mathematics Interventions. The interventions included are determined after an exhaustive search of the published and unpublished literature by the High School Mathematics review team, as well as a review of nominations submitted to the WWC. Only research on interventions that are replicable and materials based is reviewed.

Elements of Reviewable Interventions. For an intervention to be considered reliably replicable with different participants, in other settings, and at other times, it must be either

- A “branded” intervention
- or
- An unbranded intervention that meets the following conditions:
 - The intervention is described in general terms.
 - The duration of the intervention is described.
 - The characteristics of those expected to deliver the intervention are described.

A materials-based intervention should be based on text materials, manipulatives, computer software, videotapes, other materials, or any combination thereof.

Examples of possible interventions include textbooks and textbook series, software programs, and other educational technology that serves as the basis for well-defined curricula.

Types of Outcomes to be Included

Outcomes Relevant to High School Mathematics. The study needs to include at least one type of mathematics achievement measure that involves direct student assessment in at least one of the content areas. Other measures of mathematics achievement, such as student grades assigned by teachers, do not qualify as relevant outcome measures. For the High School Mathematics review, relevant outcome measures of mathematics achievement include the following:

- Standardized, nationally normed achievement tests that are appropriate for high school students
- Standardized state or local tests of mathematics achievement
- Research-based or locally developed tests or instruments that assess students' mathematical concepts or skills

Reliability of Outcome Measures. Reliability will be assessed using the following standards determined by the WWC:

- Internal consistency—minimum of 0.50
- Temporal stability/test-retest reliability—minimum of 0.40
- Inter-rater reliability—minimum of 0.50

The study must include at least one outcome measure that demonstrates sufficient reliability or face validity.

Overalignment of Outcome Measures. A study's rating will be based only on those measures (if any) that are not overaligned. Overalignment occurs with outcome measures that are more closely aligned to one of the research groups (intervention or comparison) than the other and could bias a study's results. For instance, if the outcome measure assesses mathematics achievement using some of the same materials included in the intervention (such as specific problems), it is considered to be overaligned with the intervention. In these situations, the intervention group may have an unfair advantage over the comparison group, and the effect size is not a fair indication of the intervention's effects.

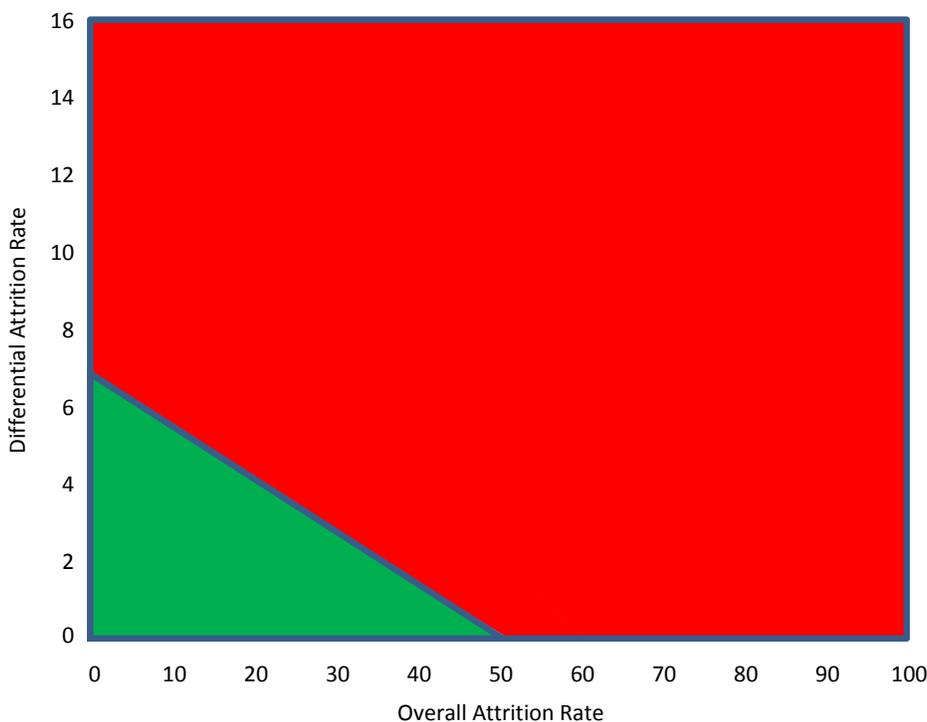
Measuring Post-Intervention Effects. A mathematics intervention may have an immediate effect as well as a longer-term effect on student mathematics achievement. Thus, outcomes measured at the end of an intervention as well as those measured any time thereafter are included. Delayed measures taken several months or years after an intervention may be useful for providing strong evidence for an intervention's longer-term effectiveness.

Types of Research to be Included

Sample Attrition. As described in the *WWC Procedures and Standards Handbook (version 2.0)*, the WWC is concerned about overall and differential attrition from the intervention and comparison groups for RCTs, as both contribute to the potential bias of the estimated effect of an

intervention. The attrition bias model developed by the WWC will be used in determining whether a study meets WWC evidence standards (see Appendix A of the *Handbook*).

When the combination of overall and differential attrition rates causes an RCT study to fall in the green area on the diagram shown below, the attrition will be considered “low” and the level of bias acceptable. However, for RCTs with combinations of overall and differential attrition rates in the red area, the attrition will be considered “high,” and these studies have potentially high levels of bias and, therefore, must demonstrate equivalence. The threshold between high and low attrition was established under the assumption that most attrition in studies of High School Mathematics is due to factors that may be slightly more related to treatment status, such as the possibility that students will change classes or subjects because of the intervention.



Many studies reviewed by the WWC are based on designs with multiple levels. Bias can be generated not only from the loss of clusters (such as schools), but also from sample members within the clusters (such as students), if those sample members attrit as a result of their treatment status. A study must pass the attrition standard at two levels. First, it must pass at the cluster level, using the attrition boundary set above. Second, the study must pass at the subcluster level, again using the attrition boundary set above, *with attrition based only on the clusters still in the sample*. That is, the denominator for the subcluster attrition calculation includes only sample members at schools or classrooms who remain in the study after cluster attrition.

Characteristics Relevant to Equating Groups. If the study design is an RCT with high levels of attrition or a QED, the study must demonstrate baseline equivalence of the intervention and

comparison groups for the analytic sample. The onus for demonstrating equivalence in these studies rests with the authors. Sufficient reporting of pre-intervention data should be included in the study report (or obtained from the study authors) to allow the review team to draw conclusions about the equivalence of the intervention and comparison groups.

Pre-intervention characteristics can include the outcome measure(s) administered prior to the intervention or other measures that are not the same as, but are highly related to, the outcome measure(s). Baseline mathematics achievement tends to be highly correlated with other characteristics that can moderate effects and, therefore, tends to be a useful measure for assessing baseline equivalence. For the High School Mathematics review, the variables on which studies must demonstrate equivalence are

- A pretest of an acceptable outcome measure
- Grade level

Groups are considered equivalent if the reported differences in pre-intervention characteristics of the groups are less than or equal to one-quarter of the pooled standard deviation in the sample, regardless of statistical significance. However, if differences are greater than 0.05 standard deviation and less than or equal to one-quarter of the pooled standard deviation in the sample, the analysis must control analytically for the individual-level pre-intervention characteristic(s) on which the groups differ. If there are pre-intervention differences greater than 0.25 for *any* of the listed characteristics, the study does not meet standards.

In addition, if there is evidence that the populations were drawn from very different settings, the principal investigator may decide that the environments are too dissimilar to provide an adequate comparison. The High School Mathematics review team also will examine other baseline characteristics (when available) to assess baseline equivalence of studies. These characteristics include, but are not limited to, the following:

- Gender
- Socioeconomic status
- Racial/ethnic breakdown
- Percentage of English as a second language students
- Percentage of bicultural students
- “At-risk” status (as provided by study authors)
- Location of the schools involved
- Homogeneous groupings of students
- School type (public, private, religious)
- School SES (e.g., Title I school)
- Average class size (small, medium, large)
- Average teacher characteristics (e.g., teacher education and experience)
- School size

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

Other Statistical and Analytical Issues. 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, the effect-size estimates will be 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 *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.

The statistical significance of group differences will be recalculated if (1) the study authors did not calculate statistical significance, (2) the study authors did not account for clustering when there was a mismatch between the unit of assignment and unit of analysis, or (3) 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 statistical significance of the effect sizes computed by the WWC will incorporate a statistical adjustment for clustering. The default intra-class correlation used for the High School Mathematics review is 0.20. For an explanation about the clustering correction, see Appendix C of the *Handbook*.

When multiple comparisons are made (i.e., multiple outcome measures are assessed within an outcome domain in one study) and not accounted for by the authors, the WWC accounts for this multiplicity by adjusting the reported statistical significance of the effect using the Benjamini-Hochberg correction. See Appendix D of the *Handbook* for the formulas the WWC uses to adjust for multiple comparisons.

All standards apply to overall findings as well as analyses of subsamples.

LITERATURE SEARCH METHODOLOGY

The literature search strategy for the WWC High School Mathematics review is two-pronged. First, the review team conducts a keyword search to identify interventions with studies that may be eligible for review. Second, the team conducts focused intervention searches to ensure that all potentially eligible studies of the identified interventions are found. Both search types are described below.

Keyword Search

The primary objective of the keyword search is to identify interventions with potentially eligible studies and assess the likely extent of studies on each intervention, so that interventions can be prioritized for review. The focus is on breadth rather than depth. The following keywords are meant to capture literature that falls within the scope of the protocol. Given the objective stated above, targeted outcomes and study design terms are included to focus the search on identifying literature that will support an intervention report. The keyword list is followed by a list of databases that are searched; the asterisk (*) in the keyword list allows the truncation of the term and will return any word that begins with the specified letters.

Achievement	Eleventh grade
Algebra	Grade 9
Analysis	Grade 10
Assessment	Grade 11
Attainment	Grade 12
Calculus	High school
Comparative	Ninth grade
Curriculum	Tenth grade
Data analysis	Twelfth grade
Fraction*	
Geometry	Alternating treatments
Inquiry-based	Assignment
Instruction	Baseline
Math*	Causal
Math* ability	Comparison group
Math* achievement	Control group
Math* aptitude	Effectiveness
Math* concept*	Evaluation
Math* instruction	Experiment
Math* skill*	Impact
Measurement	Intervention
Number*	Matched group
Operation*	Meta analysis
Outcome*	Posttest
Pattern*	Pretest
Pre-algebra	Quasi-experimental design/QED
Pre-calculus	Random
Probability	Randomized control trial/RCT
Problem solving	Regression discontinuity design
Proof	Simultaneous treatment
Reasoning	Single case design
Remedial	Single subject design
Representation	Treatment
Spatial	
Statistic*	

Supplement*
Teacher-centered
Teacher-directed
Traditional
Trigonometry

The core list of electronic databases that are searched includes the following:

- **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 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.
- **Academic Search Premier.** This multidisciplinary database provides full text for more than 4,500 journals, including 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–present. EconLit covers virtually every area related to economics.
- **Business Source Corporate.** This database 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.

- ***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.
- ***EJS E-Journals.*** E-Journals from EBSCO host® contain article-level access for thousands of E-Journals available through EBSCO's Electronic Journal Service (EJS). This resource covers journals to which Mathematica Policy Research subscribes.
- ***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.
- ***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, and abstracts and articles from academic publishers, professional societies, preprint repositories, universities, and other scholarly organizations.

In addition to the keyword search in databases, the review team seeks to identify other relevant studies through the following approaches:

- Public submissions of materials submitted via the WWC website or directly to WWC staff.
- Solicitations made to key researchers by the review team.
- Checking websites summarizing research on programs for children and youth, prior reviews, and research syntheses (i.e., using the reference lists of prior reviews and research syntheses to make sure key studies have not been omitted).
- Searches of the websites of all the developers of relevant interventions or practices for any research or implementation reports.

- Searches of the websites of more than 50 think tanks, research centers, and associations that conduct research in this topic area.

References resulting from these searches will be screened and sorted by intervention.

Intervention Search

The primary objective of the intervention search is to identify ALL effectiveness studies conducted for a specific intervention identified in the keyword search, including any that the keyword search did not identify. The strategy for the search is as follows:

- If the intervention was reviewed under the WWC Elementary School Mathematics or Middle School Mathematics reviews, rereview all references against the protocol for this topic area.
- Conduct standard library searches (searching titles and abstracts in each of the databases described above) of the intervention name.
- Scan references to identify possible synonyms for the intervention in the literature and conduct standard library searches of these terms.
- Once some potentially eligible studies are identified, request full text and review the reference lists to cross-check search results. Similarly, review relevant literature reviews. Revise search terms as needed.
- Identify seminal researchers associated with the intervention. Conduct full-text searches of the researcher name combined with the intervention name.
- Identify seminal studies of the intervention and conduct searches of the associated citations.
- Contact the intervention’s developer for a list of known research on the intervention.

All references resulting from these searches will be screened for eligibility.