WWC EVIDENCE REVIEW PROTOCOL FOR ELEMENTARY SCHOOL MATHEMATICS INTERVENTIONS VERSION 1.1[1]

Topic Area Focus

This What Works Clearinghouse (WWC) review focuses on mathematics interventions for elementary school students in grades K–5. It is focused on interventions designed to impact student achievement: both curriculum-based interventions that specify clear math learning goals for students and instructional programs and materials that organize the elementary mathematical content. 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 (that is, what students should know and be able to do) among elementary school students?
- Are some interventions more effective for certain types of students, particularly students who are at risk of failure in mathematics?

There are separate WWC reviews that focus on mathematics for older children.

Key Definitions

Math Intervention

A math intervention is defined in this review as a replicable, materials-based instructional program that

- is designed for elementary school students;
- clearly delineates math learning goals for students;
- is designed to directly affect student mathematics achievement;
- is based on text materials, computer software, videotapes, or any other materials base, or combination thereof.

[1] This protocol is aligned with the WWC Procedures and Standards Handbook (Version 1.0).

Mathematics Achievement Domain

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: numbers, arithmetic, pre-algebra, geometry, measurement, graphing, and logical reasoning. Mathematics skills are the application of the learning of this content, as well as an understanding of mathematical concepts, procedures, and problem solving. These include: problem solving, reasoning and proof, making connections, oral and written communication, and uses of mathematical representation.

GENERAL INCLUSION CRITERIA

Populations to be Included

As mentioned above, this WWC area reviews interventions for elementary school students, where "elementary school" is primarily defined as a school with any of the six grades from kindergarten through grade 5. Students in grades 6 are included in the review only if such students were classified in the study as elementary school students and were included along with students in any of the grades K through 5 in the study analysis sample. Otherwise, students in grade 6 fall within the scope of the topic area reviews for Middle School Math.

Types of Interventions to be Included

The interventions included are determined after an exhaustive search of the published and unpublished literature by the Elementary School Math review team, as well as a review of nominations submitted to the WWC. Only research on interventions that are replicable (i.e., documented well enough that they can be reproduced) and materials-based (although the materials may come in various forms) is reviewed. Examples of possible interventions to be included are as follows:

- textbooks and textbook series,
- software programs and other educational technology that serve as the basis for well-defined curricula, and
- materials including videotapes that serve as the basis for well-defined curricula.

Types of Research Studies to be Included

The elementary school math literature search focuses on studies involving math interventions for students in elementary schools (primarily defined as grades K – 5). To be included in the review, a study must meet several relevancy criteria:

Topic relevance. The study must focus on the effects of a math intervention on at least one measure of math achievement.

Timeframe relevance. The study must have been published from January 1985 to August 2008. This timeframe is wide enough to allow for a baseline of data with regard to traditional curricula, National Assessment of Educational Progress (NAEP) trends, and data with regard to standards-based curricula.

Sample relevance. The sample must include elementary school grades. Elementary school grades are defined primarily as grades K–5.

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

Study design relevance. The design must be an empirical study, using quantitative methods and inferential statistics, that includes a comparison group. Such designs include well-conducted randomized controlled trials (RCTs) and quasi-experiments with matching or equating of student samples on a baseline student-level measure of math achievement. The WWC is in the process of developing standards for reviewing and reporting on single-case design and regression discontinuity design studies; studies with these designs are not reviewed at this time, but are included in the report references with a note indicating that standards are not yet available for that research design.

Outcome relevance. The study is required to focus on student (not teacher) outcomes, and at least one of the outcome measures must be an achievement measure that demonstrates sufficient reliability or face validity.

SPECIFIC TOPIC PARAMETERS

The following parameters specify which studies of math interventions are considered for analyses and which aspects of those studies are coded for the review.

1. Elements of intervention replicability.

The important characteristics of the intervention that must be documented in a study to reliably replicate it with different participants, in other settings, at other times include the following:

- The intervention is "branded."
- The intervention is not "branded" but meets the following conditions:
  - The intervention is described in general terms.
  - The duration of the intervention is described.
  - The characteristics of the individuals who are expected to deliver the intervention are described.

2. Outcomes relevant to elementary school math.

Relevant outcomes include measures of mathematics achievement, and three types of math achievement measures are included:

- Standardized, nationally normed achievement tests that are appropriate for elementary students (e.g., Comprehensive Test of Basic Skills, Wide Range Achievement Test)
- Standardized state or local tests of math achievement
- Research-based or locally developed tests or instruments that assess students' mathematical concepts or skills

A study needs to include at least one type of math achievement measure that involves direct student assessment in at least one of the content areas. Other measures of math achievement, such as student grades assigned by teachers, do not qualify as relevant outcome measures.

The alignment between the outcome and the intervention is another factor considered in the review. Outcome measures that are more closely aligned to one of the research groups (intervention or comparison) than the other could bias a study's results. For instance, if the outcome measure assesses math achievement using some of the same materials included in the intervention (such as specific math 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. A study does not meet WWC evidence standards if it includes only one outcome measure and that measure is overaligned. Studies with more than one outcome measure will be rated based only on those measures (if any) that are not overaligned.

3. Reliability of outcome measures.

Reliability (internal consistency, temporal stability/test-retest reliability, and inter-rater 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

Studies must contain at least one relevant measure that has evidence of reliability, or uses a state or district assessment.

4. Characteristics relevant to equating groups.

Characteristics of participants that might confound a math intervention's effect and must be equated if a study is not an RCT include:

- Pretest of an acceptable outcome measure[2]
- Grade level

For a quasi-experimental design (QED) comparison study to be valid, groups of students being compared would ideally be drawn from the same population of students. Consequently, the groups would be similar along a baseline measure of math achievement. Evidence that the groups in a QED comparison study differ substantially along the pretest will result in the failure of a study because substantial differences suggest that the groups represent different populations. In particular, a study fails to meet WWC evidence standards if the pretest difference between the groups is either: (1) statistically significant and at least 0.05 standard deviations, or (2) not statistically significant but at least 0.5 standard deviations.

The onus for demonstrating initial equivalence of groups for a QED comparison study rests with the investigator. Sufficient reporting of the factors should be included (or can be obtained) to establish the initial equivalence of the groups. A QED comparison study must statistically adjust for the pretest when calculating the effect of the intervention on the outcome measure of math achievement, unless the pre-test difference between the groups is less than 0.05 standard deviations.

5. Effectiveness of the intervention across different student groups.

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

- Gender breakdown
- Baseline math achievement
- Socioeconomic status

---

[2] Baseline math 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. Nevertheless, the Elementary School Math review team also will examine other baseline characteristics (when available) to assess baseline equivalence of non-RCT studies. Other characteristics include, but are not limited to, gender, race/ethnicity, percentage of second-language students, measures of "at-risk" status, tracking level, special education, school location, and average class size. The provision of all such information, however, is not a requirement of the review.

- Racial/ethnic breakdown
- Percentage of second-language students
- Percentage of bicultural students
- "At-risk" status

Note: The definition of "at risk" is provided by the study authors. We capture the author's definition in the coding guide.

6. Effectiveness of the intervention across different settings.

An intervention's effectiveness could vary by school setting. Whether a study examines effects in different school settings does not affect the inclusion of a study for review or the rating given to the study. However, we will present in a technical appendix findings across settings of interest if the samples for settings are equivalent with regard to pretest and grade level. School settings of interest for this review include:

- Location of the schools involved
- Homogenous groupings of students
- School type (public, private, religious)
- School SES (e.g., Title I school)
- Average class size (small, medium, large)
- Average teacher characteristics—such as teacher education and experience—that could moderate effects
- School size

7. Measuring post-intervention effects.

A math intervention may have an immediate effect as well as a longer-term effect on student math 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 because they may provide strong evidence for an intervention's effectiveness.

8. Defining differential attrition for RCTs.

RCTs are examined for differential attrition, where differential attrition is defined as being greater than 5% differential loss. If differential attrition is less than or equal to 5%, we assume that the bias associated with it is minimal. If it is greater than 5%, the burden of proof shifts and the study authors need to provide evidence that differential attrition did not bias the

results. In particular, a post-attrition demonstration of group equivalence on the pretest is required.

An RCT with differential attrition fails to meet WWC evidence standards if the pretest difference between the post-attrition groups is (1) between 0.05 and 0.5 standard deviations and the study does not control for the pretest, or (2) greater than 0.5 standard deviations. An RCT with differential attrition can meet standards with reservations if (1) the pretest difference between the post-attrition groups is between 0.05 and 0.5 standard deviations and the study statistically adjusts for the pretest, or (2) the pretest difference between the post-attrition groups is less than 0.05 standard deviations.

9. Defining severe overall attrition for RCTs.

RCTs are examined for severe overall attrition, where severe overall attrition is defined as being greater than 30% loss. If overall attrition is less than or equal to 30%, we assume that the bias associated with it is minimal. If it is greater than 30%, the burden of proof shifts and the study authors need to provide evidence that overall attrition did not bias the results. In particular, a post-attrition demonstration of group equivalence on the pretest is required.

An RCT with severe overall attrition fails to meet WWC evidence standards if the pretest difference between the post-attrition groups is (1) between 0.05 and 0.5 standard deviations and the study does not control for the pretest, or (2) greater than 0.5 standard deviations. An RCT with severe overall attrition can meet standards with reservations if (1) the pretest difference between the post-attrition groups is between 0.05 and 0.5 standard deviations and the study statistically adjusts for the pretest, or (2) the pretest difference between the post-attrition groups is less than 0.05 standard deviations.

10. Statistical properties important for computing accurate effect sizes.

- For most statistics (including d-indexes), normal distribution and homogeneous variances are important properties. For correlations, there are no required desirable properties, and for odds-ratios there are no required desirable properties except the minimum of 5 observations per cell.

- In the case where a misaligned analysis is reported (i.e., unit of analysis is not the same as the unit of assignment) and the author is not able to provide a corrected analysis, the effect sizes computed by the WWC will incorporate a statistical adjustment for clustering. The default intraclass correlation used for elementary school math outcomes is 0.20. For an explanation about the clustering correction, see Appendix C of the WWC Procedures and Standards Handbook.

- In the case where multiple comparisons are made (i.e., multiple outcome measures are assessed within an outcome domain in one study), the WWC accounts for this multiplicity by adjusting the author-reported statistical significance of the effect using the Benjamini-Hochberg correction. See

Appendix D of the WWC Procedures and Standards Handbook for the formulas the WWC used to calculate statistical significance.

METHODOLOGY

Literature Search Strategies

The WWC employs comprehensive and systematic literature search strategies to identify the population of published and unpublished relevant studies. This section contains topic specific elements of the literature search (e.g., search terms, additional journals, and associations) performed prior to 2008. The final section describes an expanded search conducted in 2008.

Key Word List

Intervention
Instruction
Curriculum
Curriculum-based assessment
Curriculum evaluation
Mathematics
Number
Numbers
Operations
Algebra
Geometry
Measurement
Data analysis
Probability
Reasoning
Proof
Connections
Communication
Representation
Disposition
Attitude skills
Concepts
Problem solving
Mathematical aptitude
Mathematical skills
Mathematical concepts
Mathematics instruction
Mathematic* ability
Spatial ability
Properties mathematics
Properties
Elementary school*
Elementary grade*
Kindergarten
First grade
Second grade
Third grade
Fourth grade
Fifth grade
K-5
Outcomes of education
Mathematics achievement
Mathematics w outcome*
Outcomes
Achievement
Learn
Impact
Enhance
Supplemental math
Remedial meth
Arithmetic
Patterns
Fractions

A combination of Boolean terms such as AND and OR are used with this keyword list. WWC

staff is consulted as to the appropriate combination to use and then tailor according to each specific electronic database.

Topic-Specific List of 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. Below we list the topic-specific journals used for the elementary math hand searches.

1 Arithmetic Teacher
2 Journal for Research in Mathematics Education
3 Journal of Mathematical Behavior
4 Mathematical Thinking and Learning
5 Mathematics Teacher
6 Educational Studies in Mathematics
7 Journal of the Learning Sciences
8 Cognition and Instruction
9 Teaching Children Mathematics
10 Instructional Science
11 Learning Instruction
12 Journal of Educational Psychology

Topic-Specific List of Organizations

1. Eisenhower Clearinghouse

Topic-Specific List of Associations

1. Association for Supervision and Curriculum Development (ASCD)

2. National Council of Supervisors of Mathematics (NCSM)

3. Psychology and Mathematics Education (PME)

Personal Contacts

The Elementary School Math review team solicits studies directly from experts in the field of education who work on math interventions. The Principal Investigator (PI) identifies these experts. We also contact experts using listservs dedicated to this topic and whose members are scholars working in this area.

Direct Contact with Publishers and Developers

The National Council of Teachers of Mathematics Exhibitors' List

The National Council of Teachers of Mathematics Exhibitors' List is searched for textbook publishers and software developers who have either published or developed an elementary, middle, or secondary school math curriculum. Organizations of this type may have research results of either pretests of the curriculum or evaluations of the curriculum after it has been implemented in schools that may not appear in refereed journals. We search the websites of these exhibitors to determine if they have published or developed a curriculum-based mathematics intervention. We use the guidelines developed by the PI for determining whether the products offered by the exhibitors constitute relevant mathematics interventions. The results of this search are reviewed by the PI before the exhibitors are contacted. Initially, a standardized email is sent to the relevant publishers and developers requesting submissions of pertinent research to the What Working Clearinghouse website. A follow-up letter also is mailed to organizations that do not respond to our email request.

There is recognition that studies obtained from the publishers and developers might be biased towards a positive outcome. Articles and studies from these sources are to be flagged according to their source. However, given the few studies we have obtained so far for this math review, it is unlikely we will be able to conduct a meta-analysis and sensitivity analysis.

Note: Not all of the textbook publishers listed below are actually publishers, but they might provide information and links to textbook and curriculum series.

Textbook Publishers List

1. AAAS/Project 2061
2. Academic Systems Company
3. AGS Publishing
4. American Math Publishing
5. AMME, Inc.
6. Amsco School Publications, Inc.
7. Bates Publishing Company
8. CAROLINA Biological Supply Co.
9. Carson-Dellosa Publishing Co.
10. Charlesbridge Publishing
11. COMAP
12. Continental Press
13. CORD Communications
14. CPM Educational Program
15. Creative Teaching Press
16. Developmental Studies Center
17. Digi-Block, LLC
18. Dimension 2000
19. Education Development Center
20. Educators Publishing Service

21. ETA/Cuisenaire
22. Evan-Moor Educational Publishers
23. FASE / Foundation for Advancements in Science and Education
24. Frank Schaffer Publications
25. Glencoe/McGraw-Hill
26. Great Source Education Group
27. Holt, Rinehart & Winston
28. Holtzbrinck Publishers
29. Houghton Mifflin Harcourt Supplemental Publishers
30. J. Weston Walch, Publisher
31. Kendall/Hunt Publishing Company
32. Pearson Learning Group
33. Sadlier-Oxford
34. Summit Learning
35. Teacher Created Materials, Inc.
36. Teachers College Press
37. Tri-C Publications, Inc.
38. Venture Publishing
39. Wholemovement Geometry
40. Workman Publishing
41. Wright Group/McGraw-Hill

Software Developers List

1. Apple Computer, Inc.
2. Boxer Learning
3. Carnegie Learning, Inc.
4. Edu2000 America, Inc.
5. Learning in Motion, Inc.
6. MathMedia Educational Software, Inc.
7. Meridian Creative Group
8. New Century Education Corporation
9. Optical Data Corporation
10. Riverdeep, Inc.
11. Steck-Vaughn
12. Steck-Vaughn
13. Curriculum Associates
14. Delta Education
15. Didax, Inc.
16. EdQuest Technology
17. IntelliTools, Inc.
18. Renaissance Learning, Inc.
19. Tom Synder Productions
20. Educational Insights
21. GAMCO Educational Software
22. Neufeld Learning Systems, Inc.
23. Teacher Created Resources

24. ThinkSharp, Inc.
25. William K. Bradford Publishing Company
26. BestQuest Teaching Systems, LLC
27. Key Curriculum Press
28. The Futures Channel
29. Barnum Software

Publishers and software developers are directed to submit their studies online at the WWC website. They are also asked to call us if they have any concerns or need more information.

2008 Literature Search

In early 2008, we conducted a literature search to update the population of relevant studies published in 2005 or later. This search focused on electronic databases. This section lists the key words and databases used.

Key Word List

The list below contains the key words used during the expanded literature search for both Elementary Math and Middle School Math review areas. These terms were put into the key word list of databases with the limitation that the publication date be after January 1, 2005.

Achievement
Algebra
Arithmetic
Assessment
Attainment
Comparative math studies
Curriculum
Curriculum evaluation
Curriculum-based assessment
Curriculum evaluation
Eighth grade
Elementary schools
First grade
Fifth grade
Fourth grade
Fractions
Geometry
Grade 1
Grade 2
Grade 3
Grade 4
Grade 5
Grade 6
Grade 7
Grade 8
Grade 9
Impact
International math
Inquiry-based math
Instruction
Intervention
Kindergarten
Mathematic* ability
Mathematical aptitude
Mathematical concepts
Mathematical skills
Mathematics
Mathematics achievement
Mathematics instruction
Mathematics w outcome*
Middle grades
Middle schools
Ninth grade
Outcomes
Outcomes of education
Properties mathematics
Reasoning
Remedial math

Second grade
Seventh grade
Sixth grade
Spatial ability
Third grade

In addition to searching the above key words, we performed specific searches for each of the interventions identified in the 2007 Elementary School Math topic report.

A combination of Boolean terms such as AND and OR were used with this keyword list. Libraries at MPR conducted the actual searching and should be consulted as to the appropriate combination to use for searching within each electronic database.

List of Electronic Databases Searched

1. Academic Search Premier
2. Business Source Corporate
3. Dissertation Abstracts
4. EBSCO Ed Research Complete
5. EconLit
6. Education Research
7. EJS E-Journals
8. ERIC
9. Google Scholar
10. Organizational websites:
    a. Abt Associates (abtassociates.com)
    b. American Enterprise Institute for Public Policy Research (aei.org)
    c. Edvantia (ael.org)
    d. American Institutes for Research (air.org)
    e. Brookings Institution (brookings.edu)
    f. California Institute for Federal Policy Research (calinst.org)
    g. Center for Comprehensive School Reform and Improvement (centerforcsri.org)
    h. Center on Instruction (centeroninstruction.org)
    i. Center on Education Policy (cep-dc.org)
    j. Center for Research in Educational Policy (crep.memphis.edu)
    k. Thomas B. Fordham Institute (edexcellence.net) (The Fordham Institute has assumed the work of the Educational Excellence Network.)
    l. Thomas B. Fordham Institute (fordhaminstitute.org)
    m. Institute for Public Policy and Social Research (ippsr.msu.edu)
    n. Mid-continent Research for Education and Learning (mcrel.org)
    o. MDRC (mdrc.org)
    p. Learning Point Associates (ncrel.org)
    q. National Council of Teachers of Mathematics (nctm.org)
    r. Northwest Regional Educational Laboratory (nwrel.org)
    s. Public Policy Research Institute (ppri.tamu.edu)
    t. Public/Private Ventures (ppv.org)

   u. Pacific Resources for Education and Learning (prel.org)
   v. Public Education Network (publiceducation.org)
   w. RAND Corporation (rand.org)
   x. SEDL (sedl.org)
   y. SRI International (sri.com)
   z. Urban Institute (urban.org)
11. PolicyArchive
12. SocAbstracts
13. SocINDEX with Full Text
14. WorldCat