WWC EVIDENCE REVIEW PROTOCOL FOR ELEMENTARY SCHOOL MATHEMATICS INTERVENTIONS, VERSION 1.0

Topic Area Focus

The What Works Clearinghouse (WWC) review on elementary school math interventions focuses on curriculum-based math interventions that specify clear learning goals for students and assess student outcomes related to mathematics achievement. There are two WWC Evidence Reports planned in this topic area. One report focuses on interventions for middle school mathematics. The second report focuses on interventions for elementary school mathematics.

A systematic review of evidence in this topic area addresses the following questions:

- Which curriculum-based math interventions are effective for improving mathematics achievement outcomes for elementary school students?
- Are some math interventions more effective for improving mathematics achievement for some subgroups of students, particularly students who are at risk for failure in mathematics?

Key Definitions

Curriculum-based interventions. A curriculum-based math intervention is defined in this review as a replicable, materials-based instructional program which

- covers one or more of the following content areas: numbers, arithmetic, geometry, pre-algebra, measurement, graphing, and logical reasoning;
- specifies the population of learners as elementary school students;
- delineates the learning goals for students; and
- directly assesses student outcomes related to mathematics achievement.

Curriculum-based math interventions are considered in this review if the instructional programs:

- extend over the course of one semester or more;
- are central to students' regular school instruction; and
- are based on text materials, manipulatives, computer software, videotapes, other materials, or any combination thereof.

GENERAL INCLUSION CRITERIA

Populations to be Included

The WWC is reviewing interventions that are designed for elementary school students (i.e., children approximately 5 to 10 years of age). However, because there is some variation across school districts around the nation on how elementary school is defined and organized, for this review elementary school is primarily defined as a school with any of the six grades, K through 5. This definition of elementary school is used throughout the protocol. However, a working definition is needed since some school districts might be using other configurations. The first criterion in screening is to include studies that explicitly indicate that the schools were elementary schools or that they were using a curriculum meant for elementary schools. Otherwise, a study is included if it indicates that its sample was either kindergarten, first, second, third, fourth, or fifth grade or some combinations of these grades.

Types of Interventions to be Included

The interventions to be included are determined after an exhaustive search of the published and unpublished literature by the Evidence Report Team as well as a review of the 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) are reviewed.

Examples of possible interventions to be included are

- 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

This review includes empirical studies, published in 1985 or later, that focus on math interventions for students in elementary schools (grades K–5). The studies include students attending schools in the U.S., its territories, or its tribal entities. Research on interventions developed outside of the U.S. are considered if the studies use U.S. samples. The focus of the outcome measures is the students, not the teachers, and at least one of the outcome measures must focus on math achievement and demonstrate sufficient reliability or face validity.

The review focuses on empirical studies, using quantitative methods and inferential statistics. Specifically these include well-conducted randomized controlled trials (RCTs), quasi-experiments (QEDs) with matching or equating on the pretest, and regression discontinuity designs. At this time, the WWC has not developed standards for reviewing or reporting on single-case design studies. Consequently, studies with a single-case design are not currently

included in this review. The focus on empirical designs is reflected in the collection, review, and reporting of the research. The studies must all have adequate statistical reporting such that effect sizes can be calculated.

SPECIFIC TOPIC PARAMETERS

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

1. **The characteristics necessary to define interventions that reflect commonly shared and/or theoretically derived characteristics include:**

   1a. Characteristics of an elementary school core math curriculum:

   - Materials-based instructional program, targeted to grades K–5 or labeled as an "elementary school" curriculum
   - Covers one or more of the following content areas:
     - Numbers
     - Arithmetic
     - Geometry
     - Pre-Algebra
     - Measurement
     - Graphing
     - Logical Reasoning
   - Clearly defined learning goals (e.g., the curriculum specifies what students will know and will be able to do after its implementation)
   - Extended over the course of one semester or more of instruction
   - Enough information provided so that the intervention is replicable
   - Implemented between 1985 and 2005

   1b. We make the additional distinction between "branded" and "non-branded" interventions. Branded interventions are programs and products that may possess any of the following characteristics:

- Have an external developer who provides technical assistance (e.g., instructions/guidance on the implementation of the intervention), and who sells or distributes the intervention
- Are replicable: packaged or otherwise available for distributions/use beyond a single site
- Are trademarked

2. The variations in the intervention that are important to examine include:

- Amount of professional development
- Degree of use of text materials, manipulatives, videotapes, software, or other instructional materials specified as part of the curriculum
- Degree of use of prescribed classroom structures, such as collaborative group work
- Amount of dosage provided to students (i.e., how often the intervention was delivered to students and for how long)
- Degree of implementation fidelity
- If there is no evidence in the study that the intervention was altered when it was implemented, we will assume that it was implemented as defined.

3. The important characteristics of the intervention that we would need to know in order to reliably replicate it with different participants, in other settings, at other times include:

- The core math curriculum, as embodied in the instructional goals and materials
- The duration of the intervention (including the dosage) that is prescribed for students to master the instructional goals
- The characteristics of the individuals who are expected to deliver the intervention (e.g., teacher, aide, parent volunteer, other students)
- The dimensions of the curriculum (e.g., use of support materials, specific classroom activities) that are critical to fidelity

Branded interventions are assumed to be replicable.

4. Important classes of outcomes include measures of mathematics achievement.

Three types of outcome measures are included:

1. Standardized, nationally normed achievement tests that are appropriate for elementary students (e.g., Comprehensive Test of Basic Skills, Wide Range Achievement Test)

2. Standardized state or local tests of math achievement

3. 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.

Outcomes may include one or more of the following content areas:

- Number
- Arithmetic
- Geometry
- Pre-Algebra
- Measurement
- Graphing
- Logical Reasoning

5. Evidence sufficient for an outcome measure to demonstrate each type of reliability (internal consistency, temporal stability/test-retest, and inter-rater) includes:

Internal consistency: .60

Temporal stability: .40 or higher. Test-retest estimates can be low even when the construct is well measured (due to real change in the participants), and high even when it is measured poorly (due to memory). If reviewers find a temporal stability estimate and it is the only reliability estimate for the measure, it should be flagged and discussed with the Principal Investigator (PI).

Inter-rater reliability: .50 or higher. For inter-rater reliability, reviewers need to discern whether or not the estimate is corrected for chance. If reviewers find an inter-rater reliability estimate and it is the only reliability estimate for the measure, it should be flagged and discussed with the PI.

Studies must contain at least one relevant measure that has evidence of either face validity or reliability.

6. **The interval of time in which studies should have been conducted to be appropriate for the Evidence Report is as follows:**

   Studies need to have been conducted within the past 20 years (i.e., with a publication date of 1985 or later). This is the default time interval for all WWC reviews. This time scale is wide enough to allow for a baseline of data with regard to traditional curricula, NAEP trends, and standards-based curricula. The rule for unpublished articles is the date the document reached its final form. However, there is a recognition that a lag between published articles and unpublished articles might exist, which could cause a bias towards having more unpublished articles. This bias is unlikely to cause a significant difference because of the lengthy time interval used.

7. **The necessary characteristics that define the target population include the following:**

   Students are approximately between the ages of 5 and 10, and are enrolled in an elementary school with any grade between grades K–5.

   Students reside and attend an institution within the U.S. (including U.S. territories and tribal entities).

8. **The important characteristics of participants that might be related to the intervention's effect and must be equated if a study does not employ random assignment include the following:**

   Broadly speaking, comparison studies are valid only if student population groups are roughly equivalent with regard to:

   - Pretest of the outcome measure
   - Grade level or age

   A study employing a quasi-experimental design without a pretest (or a proxy of a pretest) of the outcome measure is excluded from the review.

   Although it is only necessary for groups to be equivalent with regard to pretest scores (and grade level), the Evidence Report Team also codes for whether the groups are equivalent along the following dimensions:

   - Gender
   - Socioeconomic status
   - Demographic and linguistic characteristics of students
   - Students at risk for failure in mathematics
   - Special education status
   - Location of the schools involved (urban, suburban, or rural)

- Average class size

9. **The characteristics of subgroups of participants that are important (a) to have variation on and (b) to test within a study to determine whether an intervention is effective within these groups include:**

    - Gender
    - Socioeconomic status
    - Racial/ethnic breakdown
    - Percentage of ELL students
    - Percentage of bicultural students
    - "At-risk" students

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

10. **The characteristics of settings that are important to test within a study to determine whether an intervention is effective within these settings include:**

    - Location of the schools involved
    - Homogenous groupings of students
    - School type (public, private, religious)
    - Average class size (small, medium, large)
    - School SES (e.g., Title 1 school)
    - School size

    The Evidence Report Team will also look for evidence that teachers are roughly equivalent with regard to:

    - Adequate training in implementing the intervention
    - Number of years of teaching experience
    - Highest level of education
    - Other teacher characteristics as defined and provided by the study author

11. The appropriate interval for measuring the intervention's effect relative to the end of the intervention is as follows:

A math intervention may have an immediate effect as well as a long-term impact on students' math achievement. Thus measures taken at the end of an intervention, as well as measures taken 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. The Evidence Report Team codes the length of the time interval between the end of an intervention and assessment of the outcome measure.

12. If students are drawn from the same local pool, which groups of individuals (e.g. students, teachers, parents, administrators) might have been able to interfere with the fidelity of the comparison if they had known who was in the intervention and comparison groups?

The students, parents, and/or teachers in either the treatment or control group could know who was in which condition (participants are not blind to condition due to the nature of the intervention).

13. For research on this topic, differential attrition from the intervention and control groups is defined as follows:

Differential attrition is defined as being greater than 7% differential loss. If differential attrition is less than or equal to 7%, we assume that the bias associated with it is minimal. If it is greater than 7%, the burden of proof shifts, and the study authors need to convince us that differential attrition did not bias the effect size estimate. A post-attrition demonstration of group equivalence on the pretest is required. We will define "post-attrition demonstration of group equivalence" as either a well-powered (.80) test of equivalence that is nonsignificant, or a standardized mean difference between groups of less than d = .10.

14. For research on this topic, severe overall attrition is defined as follows:

Severe overall attrition is defined as being greater than 20% loss. If overall attrition is less than or equal to 20%, we assume that the bias associated with it is minimal. If it is greater than 20%, the burden of proof shifts, and the study authors need to convince us that overall attrition did not bias the effect size estimate. A post-attrition demonstration of group equivalence on the pretest is required. We will define "post-attrition demonstration of group equivalence" as either a well-powered (.80) test of equivalence that is nonsignificant, or a standardized mean difference between groups of less than d = .10.

15. The statistical properties of the data that are important in order to obtain an accurate estimate of an effect size are as follows:

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

METHODOLOGY

Literature Search Strategies

The WWC Evidence Report Team 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).

Key Word List

Intervention
Instruction
Curriculum
Curriculum-based assessment
Curriculum evaluation
Mathematics
Number
Numbers
Operations
Algebra
Geometry
Measurement
Data analysis
Probability
Problem solving
Reasoning
Proof
Connections
Communication
Representation
Disposition
Attitude skills
Concepts
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 math
Arithmetic
Patterns
Fractions

A combination of Boolean terms such as AND and OR will be used with this keyword list. The librarian at AIR will be consulted as to the appropriate combination to use and then tailor according to each specific electronic database.

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.

Topic-Specific List of Journals to be Hand Searched

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 WWC Math Evidence Report Team solicits studies directly from experts in the field of education who work on math curriculum interventions. The Principal Investigators (PIs) identify 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, Developers, and Others

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 PIs for determining whether the products offered by the exhibitors constitute relevant mathematics interventions. The results of this search are reviewed by the PIs before the exhibitors are contacted. Initially, a standardized email is sent to the relevant publishers and developers requesting submissions of pertinent research to the WWC website. A follow-up letter 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 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. Harcourt Brace & Company
28. Holt, Rinehart & Winston
29. Holtzbrinck 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.