What Works Clearinghouse

ies INSTITUTE OF EDUCATION SCIENCES

Houghton Mifflin Mathematics

Program description

Houghton Mifflin Mathematics is a core curriculum for kindergarten through grade 6 students at all ability levels. According to its developer, *Houghton Mifflin Mathematics* emphasizes the five content strands and processes recommended by the National Council of Teachers of Mathematics Standards. At each grade level the program focuses on basic skills development, problem solving, and vocabulary expansion to help students master key math concepts. The program incorporates assessments—including lesson-level interventions to meet the needs of all learners—to monitor students' progress. Students practice daily math lessons through instructional software, enrichment worksheets, manipulatives, and workbooks in addition to student textbooks.

Research

Two studies of *Houghton Mifflin Mathematics* met What Works Clearinghouse (WWC) evidence standards with reservations. The two studies included students in grades 2–5 from a range of socioeconomic backgrounds, racial groups, and math performance levels. Students came from more than 800 schools in urban, suburban, and rural communities in California, Illinois, Missouri, Wisconsin, New Jersey, New York, and South Carolina.[1]

Effectiveness

Houghton Mifflin Mathematics was found to have no discernible effects on mathematics achievement.

	Mathematics achievement
Rating of effectiveness	No discernible effects
Improvement index[2]	Average: +5 percentile points Range: –5 to +12 percentile points

1. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.
2. These numbers show the average and range of improvement indices for all findings across the studies. All estimates of the improvement index across mathematics achievement outcome measures at different grade levels are positive except for a single –5 value.

Additional program information

Developer and contact

Developed by Houghton Mifflin School Division, a division of the Houghton Mifflin Company, 222 Berkeley Street, Boston, MA 02116. Web: www.hmco.com. Telephone: (617) 351-5000.

Scope of use

The edition of *Houghton Mifflin Mathematics* reviewed in this report was published in 2002. Information is not available on the number or demographics of students, schools, or districts using this program.

Teaching

Houghton Mifflin Mathematics provides three-step teaching plans for every lesson. Lessons start with an introduction or teaching step. Teachers then conduct a guided practice session, which is followed by independent student practice. The lessons end with a brief assessment and summary. Ongoing assessment is incorporated into the program to monitor students' progress. Mathematical content at each grade level is divided into topical units. The number of units for each school year ranges from 7 to 13, depending on the grade and topics covered. Each unit consists of 2–4 lessons, usually corresponding to a specific chapter in the student textbook. The developer offers professional development to school districts adopting the curriculum and additional teaching resources on its website.

Cost

Houghton Mifflin Mathematics© 2002 provides a range of individually priced classroom materials that vary by grade level and material type. Student textbooks come in single or multiple volumes and range in cost from $22.11 (grade K) to $56.46 (grades 3–6). Assessment guides for grades 1–6 are $93.24. *Houghton Mifflin Mathematics© 2002* normally provides a free teaching edition upon adoption of the curriculum. Prices for the Teacher's Edition, if purchased, range from $147.27 to $195.12, depending on grade level. Other *Houghton Mifflin Mathematics© 2002* classroom materials include individual student manipulative kits ($16.98–$19.98), teacher resource book ($33.87), homework and practice workbooks ($8.58), spiral review blackline masters ($29.46), lesson planner CD-ROM ($70.11), test preparation transparencies ($51.48), and the Test Generator CD-ROM ($121.50).

Research

Four studies reviewed by the WWC investigated the effects of the *Houghton Mifflin Mathematics* program. Two studies were quasi-experimental designs that met WWC evidence standards with reservations. The two remaining studies did not meet WWC evidence screens.

Johnson and Hall (2003) included 160 intervention schools in eight California districts using *Houghton Mifflin Mathematics* (2002 edition) in grades 2–5 and 137 comparison schools in eight different districts using non-Houghton Mifflin programs. The intervention schools had completed their first year of implementing *Houghton Mifflin Mathematics.* The comparison school districts were matched to the intervention districts based on prior mathematics achievement scores on California's Stanford 9 test, student demographic characteristics, and district sizes. Selection of comparison school districts relied on data from the California Department of Education, the Quality Education Database, and the American Institutes for Research. Statistical analyses of the math scores for the intervention districts and the comparison districts collected during the baseline year (2000–01) showed that, prior to the introduction of *Houghton Mifflin Mathematics,* there were no statistically significant differences between the two groups of schools at any grade level.

The EDSTAR, Inc. (2004) study was conducted in 519 schools from 32 school districts (16 district pairs) in California, Illinois, Missouri, Wisconsin, New Jersey, New York, and South Carolina. The intervention group included 308 schools from 16 districts using *Houghton Mifflin Mathematics* (2002 edition) for the first time during the 2002–03 school year. The comparison group included 211 schools from 16 different districts using reform, traditional, or balanced math programs. Math programs were

Research *(continued)*

classified as reform if they placed more emphasis on conceptual understanding than on traditional computation skills. Traditional programs emphasized computational skills, while balanced programs integrated conceptual understanding with traditional computational skills.

In each of the 16 district pairs in the EDSTAR, Inc. (2004) study, the intervention and comparison districts were matched based on prior mathematics achievement scores for the baseline year (2001–02), student demographics, district size, and average school size. No statistically significant differences in math achievement scores for the baseline year were found between the intervention and comparison groups. The WWC determined that having one district in the intervention group and a separate district in the comparison group confounded the intervention effect with the district.[3] The intervention effect could not be disentangled from other district characteristics without limiting the study to states that had multiple districts in the intervention and comparison groups. The authors provided additional information that enabled the district data to be separated by state. The WWC analyses are based on the reduced sample of three states, eight district pairs (16 districts), and 212 schools. The three states in the reduced sample were California, South Carolina, and New Jersey. California had two district pairs (four districts) and 68 schools. South Carolina had four district pairs (eight districts) and 128 schools. And New Jersey had two district pairs (four districts) and 16 schools. In the reduced sample all of the comparison districts within a state used the same type of math program (reform, traditional, or balanced).

Effectiveness

Findings

The WWC review of elementary school mathematics curriculum-based interventions addresses student outcomes in mathematics achievement.

Mathematics achievement. Johnson and Hall (2003) reported significant, positive effects for *Houghton Mifflin Mathematics* on overall mathematics achievement for grades 2–5. Because the authors presented average school-level math achievement gains and pretest scores but no posttest scores, the WWC requested school-level average posttest scores from the authors.[4] Using the school-level data provided by the authors, and after accounting for clustering,[5] the WWC determined that the effect of *Houghton Mifflin Mathematics* on math achievement was neither statistically significant nor substantively important, according to WWC criteria. Thus the WWC categorized the effect of *Houghton Mifflin Mathematics* on mathematics achievement as indeterminate.

The EDSTAR, Inc. (2004) study used a series of comparisons between a single treatment district and a single comparison district. This analysis does not allow the effect of the intervention to be disentangled from the effect of other district characteristics. As a result, the WWC requested that the authors aggregate the data in each of the three states that included multiple treatment districts and multiple comparison districts.[6] This

3. For more information see the WWC Technical Paper on Teacher-Intervention Confound.
4. For details about the information the WWC uses for its calculations, see Technical Details of WWC-Conducted Computations.
5. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate the statistical significance. In the case of *Houghton Mifflin Mathematics,* a correction for clustering was needed.
6. This analysis focuses on data where there was more than one district pair per state and, within the state, more than one district pair per type of comparison condition. For example, two California district pairs had comparison districts that used the balanced curriculum, so those district pairs were included in the analysis. But only one district pair in California had a comparison district using the reform curriculum, so the intervention and comparison districts in that pair were excluded from analysis. The reduced sample did not include any comparison districts using a traditional math curriculum.

Effectiveness *(continued)*

reanalysis eliminated the confound between intervention effects and district effects. Using school-level data provided to the WWC by the authors for the three states that had multiple districts in the intervention and comparison groups and accounting for clustering,[5] the WWC determined that the effect of *Houghton Mifflin Mathematics* on mathematics achievement was neither statistically significant nor substantively important, according to WWC criteria. Thus the WWC categorized the effect of *Houghton Mifflin Mathematics* on mathematics achievement as indeterminate.

Rating of effectiveness

The WWC rates the effects of an intervention in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative. The rating of effectiveness takes into account four factors: the quality of the research design, the statistical significance of the findings (as calculated by the WWC[5]), the size of the difference between participants in the intervention condition and the comparison condition, and the consistency in findings across studies (see the WWC Intervention Rating Scheme). The WWC found Houghton Miffin to have no discernible effects for mathematics achievement.

The WWC found *Houghton Mifflin Mathematics* to have no discernible effects for mathematics achievement

Improvement index

The WWC computes an improvement index for each individual finding. In addition, within each outcome domain, the WWC computes an average improvement index for each study and an average improvement index across studies (see Technical Details of WWC-Conducted Computations). The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. Unlike the rating of effectiveness, the improvement index is entirely based on the size of the effect, regardless of the statistical significance of the effect, the study design, or the analysis. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results. The average improvement index for mathematics achievement is +5 percentile points across the two studies, with a range of –5 to +12 percentile points across findings.

Summary

The WWC reviewed four studies on *Houghton Mifflin Mathematics.* Two studies met WWC evidence standards with reservations; the remaining studies did not meet WWC evidence screens. Based on these two studies, the WWC found no discernible effects on mathematics achievement. The evidence presented in this report is limited and may change as new research emerges.

References

Met WWC evidence standards with reservations

EDSTAR, Inc. (2004). *Large-scale evaluation of student achievement in districts using Houghton Mifflin.* Raleigh-Durham, NC: Author.

Additional source:

EDSTAR, Inc. (2004). *Large-scale evaluation of student achievement in districts using Houghton Mifflin Mathematics: Phase two.* Raleigh-Durham, NC: Author.

Johnson, J., & Hall, M. (2003). *Technical report: Houghton Mifflin California math performance evaluation.* Raleigh, NC: EDSTAR, Inc.

Additional source:

Johnson, J., Yanyo, L., & Hall, M. (2002). *Evaluation of student math performance in California school districts using Houghton Mifflin Mathematics.* Raleigh, NC: EDSTAR, Inc.

References *(continued)*

Did not meet WWC evidence screens

Houghton Mifflin Company. (n.d.). *Student performance in New York City District 9 on New York City/state assessments after one year of Houghton Mifflin Mathematics.* Retrieved May 4, 2006 from www.eduplace.com/state/pdf/hmm/05/efficacy/g23552_hmm05_p57-59.pdf.[7]

Mehrens, W. A., & Phillips, S. E. (1986). Detecting impacts of curricular differences in achievement test data. *Journal of Educational Measurement, 23*(3), 185–196.[7]

For more information about specific studies and WWC calculations, please see the WWC *Houghton Mifflin Mathematics* Technical Appendices.

7. Does not use a strong causal design: the study did not use a comparison group.

Appendix

Appendix A1.1 Study characteristics: Johnson & Hall, 2003 (quasi-experimental design)

Characteristic	Description
Study citation	Johnson, J., & Hall, M. (2003). *Technical report: Houghton Mifflin California math performance evaluation.* Raleigh, NC: EDSTAR, Inc.
Participants	The participants in this study were second through fifth graders from 16 districts in California. The intervention group included 160[1] schools from eight districts using *Houghton Mifflin Mathematics.* The comparison group included 137 schools in eight different districts. The intervention group was identified by Houghton Mifflin, which provided the names of eight districts in California that began using *Houghton Mifflin Mathematics* in 2002. Using data from the Quality Education Database, the California Department of Education, and the American Institutes for Research, comparison districts were matched based on prior math achievement scores, student demographic characteristics, and district sizes.
Setting	The participating school districts were located throughout California.
Intervention	The intervention group used the 2002 edition of *Houghton Mifflin Mathematics* and had completed their first year of implementing the curriculum during the 2001–2002 school year.
Comparison	There is no information in the study about the specific math programs used in the comparison school districts, except that the schools did not use *Houghton Mifflin Mathematics.*
Primary outcomes and measurement	The outcome measure was the total math score on the California statewide assessment, the Standardized and Reporting (STAR) Stanford 9 test, used during the 2000–01 and 2001–02 school years. (See Appendix A2 for more detailed descriptions of outcome measures.) The study authors reported scores as national percentile ranks, but the WWC reports scaled scores sent by the author in response to a data request, because scaled scores are more direct indicators of performance and do not require extrapolation based on national norms.
Teacher training	No information is available on the training or professional development provided to the teachers in the intervention group.

1. Some of the grade level analyses contained fewer than 160 intervention schools because not all schools had all grade levels.

Appendix A1.2 Study characteristics: EDSTAR, Inc., 2004 (quasi-experimental design)

Characteristic	Description
Study citation	EDSTAR, Inc. (2004). *Large-scale evaluation of student achievement in districts using Houghton Mifflin.* Raleigh-Durham, NC: Author.
Participants	The participating 519 schools were selected from different regions of the country including the West (California), the Midwest (Illinois, Missouri, and Wisconsin), the Northeast (New Jersey and New York), and the Southeast (South Carolina). The grade levels evaluated varied by state: California, grades 2–5; South Carolina, grades 3–5; Missouri, New Jersey, New York, and Wisconsin, grade 4; Illinois, grades 3 and 5. The authors indicate that no attrition occurred in this study. Due to the confounding of the intervention effect with the effect of other district characteristics,[1] the analysis was limited to a sample of 16 districts (eight pairs) and 212 schools in the three states that had multiple districts in the intervention and comparison groups: California, New Jersey, and South Carolina.
Setting	Districts were selected in various states to represent ranges in size, demographic characteristics, and student achievement. Within districts, schools were matched based on size of schools, student achievement level, school socioeconomic level, and school minority level.

(continued)

Appendix A1.2 Study characteristics: EDSTAR, Inc., 2004 (quasi-experimental design) *(continued)*

Characteristic	Description
Intervention	The eight districts in the intervention group had begun using *Houghton Mifflin Mathematics* in 2002–03.
Comparison	The comparison group used one of three types of math programs: reform, traditional, or balanced. The reform programs included Everyday Math, Mathland, and Excel Math. The traditional programs included Saxon and SRA. Scott Foresman 2000, Harcourt-Brace Mathematics, and Silver Burdett comprised the balanced programs. This WWC report focuses on an analysis of a reduced sample of states and therefore includes only comparison groups with balanced (California and South Carolina) and reform (New Jersey) programs.
Primary outcomes and measurement	The outcome measures were the state achievement tests used by each state in the study. Due to differences in state tests and state standards, results for each state were analyzed and evaluated separately. (See Appendix A2 for more detailed descriptions of outcome measures.) The study authors reported scores as percent of students at or above proficiency.
Teacher training	No information is available on the training or professional development provided to the teachers in the intervention group.

1. For more information see the WWC Technical Paper on Teacher-Intervention Confound.

Appendix A2 Outcome measures in the mathematics achievement domain

Outcome measure	Description
Standardized and Reporting (STAR) Stanford 9 test	Johnson and Hall (2003) used the 2001 and 2002 Stanford 9 scaled test scores to measure mathematics achievement. The test scores were obtained from the California Department of Education website.
State achievement tests	EDSTAR, Inc. (2004) used state achievement tests from California, New Jersey, and South Carolina to measure students' mathematics achievement.[1] For California, the authors used two tests from the Standardized Testing and Reporting (STAR) program of the California Assessment System: the California Standards Test and the Stanford 9 test. In 2003 the Stanford 9 test was replaced by another norm-referenced test, the California Achievement Test (as cited in EDSTAR, Inc., 2004). The California Standards Test was administered to grades 2–9 and the Stanford 9 test was administered to grades 2–11. In New Jersey, the state assessment was the Elementary School Proficiency Assessment (ESPA), which is administered to fourth-grade students. For South Carolina, the authors used results from the Palmetto Achievement Challenge Test, which was administered to students in grades 3–8.

1. Additional outcome measures (state tests for Illinois, Missouri, and Wisconsin) were reported by the study authors but are not described here because these analyses were excluded from the WWC report due to a confound between the district and the intervention.

Appendix A3 Summary of study findings included in the rating for the mathematics achievement domain[1]

Outcome measure	Study sample	Sample size (schools/ districts, except where indicated)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Houghton Mifflin Mathematics* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group[3]	WWC calculations: Mean difference[4] (*Houghton Mifflin Mathematics* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at α = 0.05)	WWC calculations: Improvement index[7]
Johnson & Hall, 2003 (quasi-experimental design)[8]								
CA STAR test: 2002 SAT9 mean scaled scores	16 California school districts: grade 2	297/16	592.52 (21.56)	586.12 (20.72)	6.40	0.30	ns	+12
CA STAR test: 2002 SAT9 mean scaled scores	16 California school districts: grade 3	296/16	618.04 (20.65)	615.11 (20.00)	2.93	0.14	ns	+6
CA STAR test: 2002 SAT9 mean scaled scores	16 California school districts: grade 4	296/16	636.87 (20.21)	632.60 (19.16)	4.27	0.22	ns	+9
CA STAR test: 2002 SAT9 mean scaled scores	16 California school districts: grade 5	293/16	657.34 (20.66)	654.13 (19.29)	3.21	0.16	ns	+7
Average[9] for mathematics achievement (Johnson & Hall, 2003)						0.21	ns	+8
EDSTAR, Inc., 2004 (quasi-experimental design)[8]								
NJ ASK4 exam: percent at or above proficiency, 2002–03	New Jersey: grade 4	16/4	40.50 (22.00)	37.70 (21.90)	2.80	0.13	ns	+5
SC PACT exam: percent at or above proficiency, 2002–03	South Carolina: grades 3–5	128/8	34.30 (15.20)	32.10 (13.10)	2.20	0.15	ns	+6
CA CAT/6 exam: percent at or above proficiency, 2002–03	California: grades 2–5	68/4	36.40 (18.30)	38.70 (16.60)	–2.30	–0.13	ns	–5
Average[9] for mathematics achievement (EDSTAR,Inc., 2004)						0.05	ns	+2
Domain average[9] for mathematics achievement across all studies						0.13	na	+5

ns = not statistically significant
na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The intervention and control group values are based on information provided by the authors for both the Johnson and Hall (2003) and EDSTAR, Inc. (2004) studies. These values may differ from what appeared in the original studies.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.

(continued)

Appendix A3 Summary of study findings included in the rating for the mathematics achievement domain *(continued)*

7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Johnson and Hall (2003) and EDSTAR, Inc. (2004), a correction for clustering was needed, so the statistical significance reported by the WWC may differ from that reported by the study authors.
9. The WWC-computed average effect size for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.

Appendix A4 *Houghton Mifflin Mathematics* rating for the mathematics achievement domain

The WWC rates the effects of an intervention in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of mathematics achievement, the WWC rated *Houghton Mifflin Mathematics* as having no discernible effects. It did not meet the criteria for positive effects because no studies met WWC evidence standards for a strong design or showed significant, positive effects. Further, it did not meet the criteria for other ratings (potentially positive, mixed, potentially negative, and negative effects) because neither of the two studies showed statistically significant or substantively important effects, either positive or negative.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative*.

  Met. The two studies of *Houghton Mifflin Mathematics* showed indeterminate effects.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. The WWC analysis found no statistically significant positive effects in this domain.

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

  Met. The WWC analysis found no statistically significant or substantively important negative effects in this domain.

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

  Not met. The WWC analysis found no statistically significant or substantively important positive effects in this domain.

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect. Fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Not met. Two studies showed indeterminate effects, and no studies of *Houghton Mifflin Mathematics* showed statistically significant or substantively important effects, either positive or negative.

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect. At least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

  Not met. The WWC analysis found no statistically significant or substantively important effects in this domain.

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

  Not met. The WWC analysis found no statistically significant or substantively important effects in this domain.

(continued)

Appendix A4 *Houghton Mifflin Mathematics* rating for the mathematics achievement domain *(continued)*

Potentially negative effects: Evidence of a negative effect with no overriding contrary evidence

- Criterion 1: At least one study showing a statistically significant or substantively important *negative* effect.

  Not met. The WWC analysis found no statistically significant or substantively important negative effects in this domain.

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

  Met. The WWC analysis found no statistically significant or substantively important positive effects in this domain.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *negative* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. The WWC analysis found no statistically significant or substantively important negative effects in this domain.

- Criterion 2: No studies showing statistically significant or substantively important *positive* effects.

  Met. The WWC analysis found no statistically significant or substantively important positive effects in this domain.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain level effects. The WWC also considers the size of the domain level effects for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.