# Appendix

## Appendix A1.1 Study characteristics: Carroll, 1998 (quasi-experimental design)

| Characteristic | Description |
|---|---|
| **Study citation** | Carroll, W. M. (1998). Geometric knowledge of middle school students in a reform-based mathematics curriculum. *School Science and Mathematics,* 98(4), 188–197. |
| **Participants** | The participants in this study were fifth graders. The study also included sixth graders, but that grade level is not within the scope of this review. Four classes of fifth graders from four districts that had been using *Everyday Mathematics* since kindergarten were selected as the intervention group, and four classes of fifth graders from similar districts that had been using basal mathematics texts were selected as the comparison group. All classes included students of mixed ability. Only students who took both the pretest and posttest were included in the analyses. The final sample consisted of 76 students in the intervention group and 91 students in the comparison group. |
| **Setting** | The study author indicates that the participating school districts ranged from urban to rural to suburban and included students from a wide range of social and ethnic backgrounds. |
| **Intervention** | All students that participated had been using the *Everyday Mathematics* curriculum since kindergarten, so the districts had been implementing *Everyday Mathematics* for at least five years. |
| **Comparison** | The author describes the comparison group as students that had used more traditional basal mathematics texts at all previous grade levels. |
| **Primary outcomes and measurement** | Researcher-developed assessment of geometric knowledge consisting of 21 questions based on the van Hiele model of five levels of geometric understanding. (See Appendix A2 for more detailed descriptions of outcome measures.) |
| **Teacher training** | Teachers were provided with instructions for administering the test. No teacher training in the use of the curriculum was reported. |

## Appendix A1.2 Study characteristics: Riordan & Noyce, 2001 (quasi-experimental design)

| Characteristic | Description |
|---|---|
| **Study citation** | Riordan, J., & Noyce, P. (2001). The impact of two standards-based mathematics curricula on student achievement in Massachusetts. *Journal for Research in Mathematics Education,* 32(4), 368–398. |
| **Participants** | The participants in this study were fourth-graders. A total of 67 schools were identified as using *Everyday Mathematics.* Seventy-eight comparison schools were matched on baseline mean school performance on the previous statewide mathematics test, percentage of students receiving free or reduced lunch, ethnicity, and percentage of students who had limited English language proficiency and required special education services. The final sample consisted of 3,781 students in the intervention group and 5,012 students in the comparison group. |
| **Setting** | All schools were located in Massachusetts. Overall, schools in this study had a higher percentage of white students (around 90%) and a lower percentage of students eligible for free or reduced lunch (around 10%) when compared with the state average. Also, intervention and comparison schools had performed above the state mean on statewide achievement tests. |
| **Intervention** | The 67 schools in the intervention group had implemented *Everyday Mathematics* for at least two years by 1999. Forty-eight schools in the intervention group had implemented *Everyday Mathematics* for four or more years (early implementers) and 19 schools had implemented the curriculum for two or three years (later implementers). |
| **Comparison** | The 78 matched comparison schools used 15 different textbook programs that, in aggregate, represented the instructional norm in Massachusetts. The most commonly used programs were published by Addison-Wesley, Houghton-Mifflin, and Scott Foresman. |
| **Primary outcomes and measurement** | Massachusetts Comprehensive Assessment System, a criterion-referenced state test that includes both multiple-choice and open-response questions. (See Appendix A2 for more detailed descriptions of outcome measures.) |
| **Teacher training** | None reported. |

## Appendix A1.3 Study characteristics: Waite, 2000 (quasi-experimental design)

| Characteristic | Description |
|---|---|
| **Study citation** | Waite, R. (2000). A study of the effects of *Everyday Mathematics* on student achievement of third-, fourth-, and fifth- grade students in a large North Texas Urban School District. Unpublished doctoral dissertation, University of North Texas, Denton. |
| **Participants** | The participants were third-, fourth-, and fifth-grade students. Six schools that were in their first year of implementing *Everyday Mathematics* volunteered to participate in this study, and a comparison group of 12 schools in the same school district were selected and matched on previous mathematics scores, socioeconomic status, and ethnicity. The final sample consisted of 732 students in the intervention group and 2,704 students in the comparison group. |
| **Setting** | All the schools in this study were located in a large urban school district in north Texas. |
| **Intervention** | The intervention group consisted of six schools that were part of a pilot program and volunteered to participate in this study. The intervention schools were in their first year of implementing *Everyday Mathematics* in the 1998–1999 school year. |
| **Comparison** | Based on a profile of the intervention group, a comparison group of 12 schools in the same district that were similar in socioeconomic status, grade level, ethnic diversity, and previous year's Iowa Test of Basic Skills mathematics score were selected. The comparison group used a more traditional mathematics curriculum approved by the school district. |
| **Primary outcomes and measurement** | 1999 Texas Assessment of Academic Skills mathematics scores. (See Appendix A2 for more detailed descriptions of outcome measures.) |
| **Teacher training** | Teachers in the intervention schools received 40 hours of training for the use of the *Everyday Mathematics* curriculum and also received the "Teacher's Resource Package." |

## Appendix A1.4 Study characteristics: Woodward & Baxter, 1997 (quasi-experimental design)

| Characteristic | Description |
|---|---|
| **Study citation** | Woodward, J., & Baxter, J. (1997). The effects of an innovative approach to mathematics on academically low achieving students in inclusive settings. *Exceptional Children,* 63(3), 373–388. |
| **Participants** | The participants in this study were third graders. Five classes of third graders in two schools that had been using *Everyday Mathematics* were selected as the intervention group, and four classes of third graders in one similar school, matched on student demographics and geographical location, were selected as the comparison group. All classes included students of mixed ability. The final sample consisted of 104 students in the intervention group and 101 students in the comparison group. |
| **Setting** | The three schools were located in the Pacific Northwest of the United States. They were all middle-class, suburban elementary schools and had very low percentages of students on free or reduced lunch. |
| **Intervention** | The intervention group consisted of five classes in two schools that were using *Everyday Mathematics.* They were in the third year of implementing the *Everyday Mathematics* curriculum. The intervention group consisted of 16 low-ability students, 27 average-ability students, and 61 high-ability students. |
| **Comparison** | The comparison group was selected from one school that used Heath Mathematics as their core curriculum, a more traditional approach focusing on computational skills. The comparison group consisted of 22 low-ability students, 42 average-ability students, and 37 high-ability students. |
| **Primary outcomes and measurement** | 1994 Iowa Test of Basic Skills.[1] (See Appendix A2 for more detailed descriptions of outcome measures.) |
| **Teacher training** | None reported |

1. The study also reported outcomes on an Informal Math Assessment that assessed problem solving, not overall math achievement. Since this measure was administered to a small subsample of students and was scored subjectively according to a 5-point rubric, it did not meet WWC standards and, therefore, was not included in this report.

## Appendix A2 Outcome measures in the math achievement domain

| Outcome measure | Description |
|---|---|
| **Iowa Test of Basic Skills (ITBS)** | Woodward & Baxter (1997) used one standardized measure of math achievement study. The third (Form G) of the Iowa Test of Basic Skills (ITBS) was used as both a pretest and posttest. This norm-referenced test has well documented reliability and validity. |
| **Massachusetts Comprehensive Assessment System (MCAS)** | As cited in Riordan & Noyce (2001), the Massachusetts Comprehensive Assessment System is administered annually and covers four strands of mathematics: number sense; patterns, relations, and functions; geometry and measurement; and statistics and probability. Each strand contributes at least 20% of total points and is tested with open-response, short-answer, and multiple-choice items. Raw scores are converted from scaled scores that range from 200–280. Reliability is estimated at 0.87 for grade 4. |
| **Researcher-developed assessment of geometric knowledge** | As cited in Carroll (1998), the van Hiele model for geometric understanding was used as a framework for constructing the pretest and posttest assessments. The pretest and posttest consisted of 21 questions, seven from each of the first three van Hiele levels of geometric reasoning. The authors indicated that the pretest was piloted on a smaller group of students the previous year and that it was reviewed by three mathematics researchers outside of the project. This outcome measure was determined to have face validity. |
| **1999 Texas Assessment of Academic Skills** | As cited in Waite (2000), the 1999 Texas Assessment of Academic Skills was a criterion-referenced assessment, developed by the Texas Education Agency (TEA) from the state-mandated curriculum to assess higher order thinking and problem-solving skills across all public schools in Texas. TEA reports an internal consistency reliability range of 0.88 to 0.92 for the assessment. Only the mathematics score from this assessment was used in this study. |

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

| | | | Author's findings from the study | | WWC calculations | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Mean outcome (standard deviation[2]) | | | | | |
| Outcome measure | Study sample | Sample size (students/ schools) | *Everyday Mathematics* group (column 1) | Comparison group (column 2) | Mean difference[3] (column 1– column 2) | Effect size[4] | Statistical significance[5] (at $\alpha = 0.05$) | Improvement index[6] |
| **Carroll, 1998 (quasi-experimental design)** | | | | | | | | |
| A 21-item researcher developed geometry test | Fifth graders in four schools | 167/8 | 11.9[7] (5.3) | 10.2 (4.0) | 1.70 | 0.37 | ns | +14 |
| **Average[8] for math achievement (Carroll, 1998)** | | | | | | 0.37 | ns | +14 |
| **Riordan & Noyce, 2001 (quasi-experimental design)** | | | | | | | | |
| MCAS mathematics test 1999 | Grade 4 (early implementer schools) | 6,009/99 | 248.27 (nr) | 243.11 (nr) | 5.16 | na[9] | Statistically significant | na[9] |
| **Average[8] for math achievement (Riordan & Noyce, 2001, early implementers)** | | | | | | na[9] | Statistically significant | na[9] |
| MCAS mathematics test 1999 | Grade 4 (later implementer schools) | 2,784/46 | 241.57 (nr) | 238.59 (nr) | 2.98 | na[9] | ns | na[9] |
| **Average[8] for math achievement (Riordan & Noyce, 2001, later implementers)** | | | | | | na[9] | ns | na[9] |
| **Waite, 2000 (quasi-experimental design)** | | | | | | | | |
| Texas Assessment of Academic Skills mathematics test | Grades 3, 4, and 5 | 3,346/18 | 78.82 (11.5) | 74.93 (14.8) | 3.89 | 0.27 | ns | +11 |
| **Average[8] for math achievement (Waite, 2000)** | | | | | | 0.27 | ns | +11 |

*(continued)*

## Appendix A3 Summary of study findings included in the rating for the math achievement domain[1] *(continued)*

| | | | Author's findings from the study | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Mean outcome (standard deviation[2]) | | WWC calculations | | | |
| Outcome measure | Study sample | Sample size (students/schools) | *Everyday Mathematics* group (column 1) | Comparison group (column 2) | Mean difference[3] (column 1–column 2) | Effect size[4] | Statistical significance[5] (at $\alpha = 0.05$) | Improvement index[6] |
| Woodward & Baxter, 1997 (quasi-experimental design) | | | | | | | | |
| Iowa Test of Basic Skills mathematics test | Grade 3 | 205/3 | 59.47[7] (11.9) | 61.48 (11.4) | –2.01 | –0.17 | ns | –7 |
| **Average[8] for math achievement (Woodward & Baxter, 1997)** | | | | | | –0.17 | ns | –7 |
| **Domain average[8] for math achievement across all studies** | | | | | | 0.16 | na | +6 |

ns = not statistically significant
na = not applicable
nr = not reported

1. This appendix reports findings considered for the effectiveness rating and the improvement index. Subtest findings from the same studies are not included in these ratings, but are reported in Appendix A4.
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. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, please see the Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups. The level of statistical significance was calculated by the WWC and corrects for clustering within classrooms or schools and for multiple comparisons. For an explanation see the WWC Tutorial on Mismatch. See the Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of the *Everyday Mathematics* report, a correction for clustering was needed.
6. 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.
7. The WWC reports different means than the study authors because the WWC took into account the pretest difference between the study groups. In this table, the *Everyday Mathematics* group mean equals the comparison group mean plus the mean difference.
8. The WWC-computed average effect sizes 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.
9. Student-level standard deviations were not available for this study. School-level standard deviations for early implementers were 7.9 for the intervention group and 7.2 for the comparison group. School-level standard deviations for later implementers were 8.1 for the intervention group and 6.2 for the comparison group. Because the student-level effect size and improvement index could not be computed, the magnitude of the effect size was not considered for rating purposes. However, the statistical significance for this study is comparable to other studies and is included in the intervention rating. For further details, please see Technical Details of WWC-Conducted Computations.

## Appendix A4 Summary of subtest findings in the math achievement domain[1]

| | | | Author's findings from the study | | WWC calculations | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Mean outcome (standard deviation[2]) | | | | | |
| Outcome measure | Study sample | Sample size (students/schools) | *Everyday Mathematics* group (column 1) | Comparison group (column 2) | Mean difference[3] (column 1–column 2) | Effect size[4] | Statistical significance[5] (at $\alpha = 0.05$) | Improvement index[6] |
| Waite, 2000 (quasi-experimental design) | | | | | | | | |
| TAAS math: concepts | Grades 3, 4, and 5 | 3,346/18 | 17.51 (2.6) | 16.75 (3.1) | 0.76 | 0.25 | ns | +10 |
| TAAS math: operations | Grades 3, 4, and 5 | 3,346/18 | 13.08 (2.9) | 12.2 (3.5) | 0.88 | 0.26 | ns | +10 |
| TAAS math: problem solving | Grades 3, 4, and 5 | 3,346/18 | 9.73 (3.6) | 8.63 (3.6) | 1.10 | 0.31 | ns | +12 |
| Woodward & Baxter, 1997 (quasi-experimental design) | | | | | | | | |
| ITBS math: computations | Grade 3 | 205/3 | 24.10[7] (4.7) | 27.02 (4.8) | –2.92 | –0.61 | ns | –23 |
| ITBS math: concepts | Grade 3 | 205/3 | 20.59[7] (4.5) | 18.9 (4.4) | 1.69 | 0.38 | ns | +15 |
| ITBS math: problem solving | Grade 3 | 205/3 | 14.78[7] (4.7) | 15.55 (4.2) | –0.77 | –0.17 | ns | –7 |

ns = not statistically significant

1. This appendix presents subtest findings from two measures of math achievement. It was determined that the subtests from these mathematics measures met WWC criterion for reliability or validity. The intervention rating was based on total test scores, which are presented in Appendix A3.
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. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, please see the Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The level of statistical significance was calculated by the WWC and corrects for clustering within classrooms or schools. For an explanation see the WWC Tutorial on Mismatch. See the Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of the *Everyday Mathematics* report, a correction for clustering was needed.
6. 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.
7. The WWC reports different means than the study authors because the WWC took into account the pretest difference between the study groups. In this table, the *Everyday Mathematics* group mean equals the comparison group mean plus the mean difference.

## Appendix A5 Rating for the math achievement domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of math achievement, the WWC rated *Everyday Mathematics* as having potentially positive effects. It did not meet the criteria for positive effects, because no *Everyday Mathematics* studies met WWC evidence standards for a strong design. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered, because *Everyday Mathematics* was assigned the highest applicable rating.

### Rating received

**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, thus qualifying as a *positive* effect.

  **Met.** Two studies showed substantively important positive effects. A third study showed a statistically significant positive effect.

- 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.

  **Met.** The WWC analysis found no statistically or substantively important negative effect. One study showed an indeterminate effect and three studies showed substantively important positive 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 studies that met WWC evidence standards for a strong design.

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

  **Met.** The WWC analysis found no significantly significant or substantively important negative effects.

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

## Appendix A6 Extent of evidence by domain

| | | Sample size | | |
|---|---|---|---|---|
| Outcome domain | Number of studies | Schools | Students | Extent of evidence[1] |
| **Math achievement** | 4 | 174 | 12,511 | Moderate to large |

1. A rating of "moderate to large" requires at least two studies and two schools across studies in one domain and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is "small."