What Works Clearinghouse™

INSTITUTE OF EDUCATION SCIENCES

Elementary School Mathematics | Updated February 2013

Investigations in Number, Data, and Space®

Report Contents

Program Description[1]

Investigations in Number, Data, and Space®, published by Pearson Scott Foresman, is an activity-based K–5 mathematics curriculum designed to help students understand number and operations, geometry, data, measurement, and early algebra. Each instructional unit focuses on a particular content area and lasts from two to five-and-a-half weeks. The curriculum encourages students to develop their own strategies for solving problems and engage in discussion about their reasoning and ideas. Students work in a variety of situations, including as individuals, in pairs, in small groups, and as part of the whole class.

Research[2]

The What Works Clearinghouse (WWC) identified two studies of *Investigations in Number, Data, and Space*® that both fall within the scope of the Elementary School Mathematics topic area and meet WWC evidence standards. One study meets WWC evidence standards without reservations, and one study meets WWC evidence standards with reservations, and together, they include more than 8,000 students in grades 1–2 and grades 4–5 in 16 districts across 13 states.[3] One of the studies examined math achievement after students experienced the curriculum for two years, while the other examined students after one year of curriculum experience.

The WWC considers the extent of evidence for *Investigations in Number, Data, and Space*® on the math performance of elementary school students to be medium to large for one outcome domain—mathematics achievement—examined for studies reviewed under the Elementary School Mathematics topic area.

Effectiveness

Investigations in Number, Data, and Space® was found to have potentially positive effects on mathematics achievement for elementary school students.

Table 1. Summary of findings[4]

		Improvement index *(percentile points)*				
Outcome domain	**Rating of effectiveness**	**Average**	**Range**	**Number of studies**	**Number of students**	**Extent of evidence**
Mathematics achievement	Potentially positive effects	+2	–4 to +10	2	8,393	Medium to large

Program Information

Background

Investigations in Number, Data, and Space® was developed by Technical Education Research Centers (TERC) in Cambridge, MA, and is distributed by Pearson Scott Foresman, a division of Pearson Education, Inc. Address: One Lake St., Upper Saddle River, NJ 07458. Email: communications@pearsoned.com. Web: http://www.pearsoned.com. Telephone: (201) 236-7000.

Program details

The *Investigations in Number, Data, and Space*® curriculum is organized into units within each grade: the kindergarten program contains seven instructional units, and grades 1–5 each have nine units. Each unit lasts from two to five-and-a-half weeks and is designed to be taught in sequence, building on one another. Kindergarten students receive 40–60 minutes of daily mathematics instruction, including 10–15 minutes spent on work that occurs outside of the math lesson—this additional work includes activities to practice and review key concepts that support the regular math work.

Students in grades 1–5 receive 70–75 minutes of daily mathematics instruction, including 10–15 minutes spent on additional work. Sessions include one or more of four types of activities: (a) math activities, during which students engage in hands-on activities intended to improve math skills; (b) whole-class discussions, during which students compare methods, results, and conclusions; (c) math workshop, in which students work individually, in pairs, or small groups; and (d) assessments, during which students are assessed through either written activities or observations. Follow-up for each session may consist of homework using cards or the student handbook. Teachers may also send home letters that introduce families to the concepts in each unit and provide suggestions for related activities to try at home.

Cost[5]

Investigations in Number, Data, and Space® materials can be purchased as individual instructional units or as a core curriculum package that includes all units for a classroom, a teacher's edition for each unit, the *Implementing Investigations* guide, and other program resources. For kindergarten, the core curriculum package costs $350.47 per classroom, the teacher's edition for individual units cost $49.97, and student activity books for all units cost $15.47 for a three-year license and $28.97 for a six-year license. For grades 1–5, the core curriculum package costs $445.97 per classroom. Student math handbooks cost $17.47 per student, and consumable student activity books cost $19.47 per student. Digital versions of the student activity books are available for all grades. Teacher editions and student activity books can be purchased by individual unit, with teacher editions at a cost of $49.97 per unit per classroom and student activity books at a cost of $4.20 per unit per student. For grades K–5, a core curriculum package with a manipulative kit and/or interactive whiteboard also is available, and ranges from $431.97 to $1,184.47 depending on the grade and which additional components (the manipulatives and/or whiteboard) are selected. Other ancillary materials for grades K–5, such as individual line items, overhead manipulative items, and individual card packages, can be purchased separately.

Research Summary

The WWC identified 44 studies that investigated the effects of *Investigations in Number, Data, and Space®* on the *mathematics achievement* for elementary school students.

Table 2. Scope of reviewed research[6]

Grade	1, 2, 4, 5
Delivery method	Whole class
Program type	Curriculum

The WWC reviewed eight of those studies against group design evidence standards. One study (Agodini, Harris, Thomas, Murphy, & Gallagher, 2010) is a randomized controlled trial that meets WWC evidence standards without reservations, and one study (Gatti & Giordano, 2010) is a randomized controlled trial that meets WWC evidence standards with reservations. These two studies are summarized in this report. Six studies do not meet WWC evidence standards. The remaining 36 studies do not meet WWC eligibility screens for review in this topic area. Citations for all 44 studies are in the References section, which begins on p. 5.

Summary of study meeting WWC evidence standards without reservations

Agodini et al. (2010) presented results for 110 elementary schools that had been randomly assigned to one of four conditions: *Investigations in Number, Data, and Space®* (28 schools), *Math Expressions* (27 schools), *Saxon Math* (26 schools), and *Scott Foresman–Addison Wesley Elementary Mathematics* (29 schools). The analysis included 4,716 first-grade students and 3,344 second-grade students who were evenly divided among the four conditions. The study compared average spring math achievement of students in each condition. The study reported student outcomes for both grade levels after one school year of program implementation. Student outcomes were measured by the Early Childhood Longitudinal Study–Kindergarten (ECLS-K) math assessment.

Summary of study meeting WWC evidence standards with reservations

Gatti and Giordano (2010) conducted a 2-year randomized controlled trial with a cohort of first-grade students and a cohort of fourth-grade students from eight schools across four states. In the first year of the study, teachers were randomly assigned to either the *Investigations in Number, Data, and Space®* curriculum or the regular math curricula. Students were randomly assigned to the classrooms by either school administrators or district administrators, with a few exceptions to accommodate parent requests and student needs. In the second year, the schools tried to keep these students in their randomly assigned condition. However, teachers were allowed to select the curriculum they wanted to teach, with the requirement that at least one teacher in each school would implement the *Investigations in Number, Data, and Space®* program, and at least one teacher would offer the regular curriculum.[7] The authors reported outcome data for only the 333 students who remained in their randomly assigned condition for the two years and completed the baseline assessment. Despite high attrition, the difference between the intervention and comparison groups along baseline math achievement was in the range where the study could meet WWC evidence standards with reservations, provided the results were adjusted for the baseline differences. The authors made this adjustment, and therefore, the study meets WWC evidence standards with reservations. To measure mathematics achievement, students were administered the Group Mathematics Assessment and Diagnostic Evaluation (GMADE) in the first and last month of each study year.

Effectiveness Summary

The WWC review of *Investigations in Number, Data, and Space*® for the Elementary School Mathematics topic includes student outcomes in one domain: mathematics achievement. The findings below present the authors' estimates and WWC-calculated estimates of the size and statistical significance of the effects of *Investigations in Number, Data, and Space*® on elementary school students. For a more detailed description of the rating of effectiveness and extent of evidence criteria, see the WWC Rating Criteria on p. 19.

Summary of effectiveness for the mathematics achievement domain

Two studies that meet WWC standards with or without reservations reported findings in the mathematics achievement domain.

Agodini et al. (2010) measured program impacts after one year of study participation and reported a statistically significant negative difference in mathematics achievement between first-grade students in the *Investigations in Number, Data, and Space*® group and students in one of the three comparison groups, *Math Expressions*, on the ECLS-K math assessment. However, when this result was adjusted for multiple comparisons, the authors found, and the WWC confirmed, that this difference was no longer statistically significant. Also, there were no statistically significant differences in mathematics achievement between the *Investigations in Number, Data, and Space*® first-grade group and the other two comparison groups, *Saxon Math* and *Scott Foresman–Addison Wesley Elementary Math*. For second-grade students who also participated in the study for one year, no statistically significant differences in mathematics achievement were found between the *Investigations in Number, Data, and Space*® group and the three comparison groups. The average effect size differences between the curriculum groups in both grades (first and second) were not large enough to be considered substantively important according to WWC criteria (i.e., an effect size of at least 0.25). The WWC characterizes these study findings as an indeterminate effect.

Gatti and Giordano (2010) measured program impacts after two years of study participation and reported a statistically significant positive difference in mathematics achievement between the fourth-grade cohort students in the *Investigations in Number, Data, and Space*® group and the comparison group on the GMADE assessment.[8,9] The WWC confirmed that this finding was statistically significant after adjusting for multiple comparisons. No statistically significant differences in mathematics achievement were found between the first-grade cohort students in the *Investigations in Number, Data, and Space*® group and the comparison group on the GMADE assessment. The average effect size difference between the curriculum groups for the first-grade cohort was not large enough to be considered substantively important according to WWC criteria. The WWC characterizes these study findings as a statistically significant positive effect.

Thus, for the mathematics achievement domain, one study showed an indeterminate effect and one study showed a statistically significant positive effect. This results in a rating of potentially positive effects, with a medium to large extent of evidence.

Table 3. Rating of effectiveness and extent of evidence for the mathematics achievement domain

Rating of effectiveness	Criteria met
Potentially positive effects *Evidence of a positive effect with no overriding contrary evidence.*	In the two studies that reported findings, the estimated impact of the intervention on outcomes in the *mathematics achievement* domain was one study that showed an indeterminate effect and one study that showed a statistically significant positive effect.
Extent of evidence	**Criteria met**
Medium to large	Two studies that included 8,393 students in 118 schools reported evidence of effectiveness in the *mathematics achievement* domain.

References

Study that meets WWC evidence standards without reservations

Agodini, R., Harris, B., Thomas, M., Murphy, R., & Gallagher, L. (2010). *Achievement effects of four early elementary school math curricula: Findings for first and second graders* (NCEE 2011-4001). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/pubs/20114001/pdf/20114001.pdf

Additional source:

Agodini, R., & Harris, B. (2010). An experimental evaluation of four elementary school math curricula. *Journal of Research on Educational Effectiveness, 3*(3), 199–253.

Agodini, R., Harris, B., Atkins-Burnett, S., Heaviside, S., Novak, T., & Murphy, R. (2009). *Achievement effects of four early elementary school math curricula: Findings from first graders in 39 schools* (NCEE 2009-4052). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Agodini, R., Harris, B., Atkins-Burnett, S., Heaviside, S., Novak, T., & Murphy, R. (2009). *Achievement effects of four early elementary school math curricula: Findings from first graders in 39 schools* (NCEE 2009-4053). Executive Summary. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Study that meets WWC evidence standards with reservations

Gatti, G., & Giordano, K. (2010). *Pearson Investigations in Number, Data, & Space efficacy study: Final report.* Pittsburgh, PA: Gatti Evaluation, Inc.

Studies that do not meet WWC evidence standards

Budak, A. (2009). The impact of a standards-based curriculum and teaching practices on students' mathematics achievement. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 69*(10-A), 3841. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Goodrow, A. (1998). Children's construction of number sense in traditional, constructivist, and mixed classrooms. *Dissertation Abstracts International, 59*(04), 1055A. (UMI No. 9828874) The study does not meet WWC evidence standards because the measures of effectiveness cannot be attributed solely to the intervention – there was only one unit assigned to one or both conditions.

Heinerikson, L. (2006). *The effects of Scott Foresman's mathematical Investigations curriculum on elementary standardized test scores* (Unpublished master's thesis). Northwest Missouri State University, Maryville. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Kehle, P., Essex, K., Lambdin, K., & McCormick, K. (2007). *What did they learn? A longitudinal, comparative, and focused study of a prepublication version of Investigations in Number, Data, and Space.* Evaluation report. Columbus, IN: Indiana University. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

McCormick, K. K. (2006). Examining the relationship between a standards-based elementary mathematics curriculum and issues of equity. *Dissertation Abstracts International, 66*(08A), 2872. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Additional source:

McCormick, K. K. (2005). *Third-grade students, a standards-based mathematics curriculum, and issues of equity.* Paper presented at the 27th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education (PME-NA), Roanoke, VA.

Mokros, J., Berle-Carmen, M., Rubin, A., & Wright, T. (1994). *Full year pilot grades 3 and 4: Investigations in Number, Data, and Space.* Retrieved from http://investigations.terc.edu/impact/1stEd/pilot3-4.cfm The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Studies that are ineligible for review using the Elementary School Mathematics Evidence Review Protocol

Battista, M. T., & Clements, D. H. (1996). Students' understanding of three-dimensional rectangular arrays of cubes. *Journal for Research in Mathematics Education, 27*(3), 258–292. The study is ineligible for review because it does not examine the effectiveness of an intervention.

Additional source:

Battista, M. T., & Clements, D. H. (1998). Students' understanding of three-dimensional cube arrays: Findings from a research and curriculum development project. In D. Chazan & R. Lehrer (Eds.), *Designing learning environments for developing understanding of geometry and space* (pp. 227–248). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Bay-Williams, J. M., Scott, M. B., & Hancock, M. (2007). Case of the mathematics team: Implementing a team model for simultaneous renewal. *Journal of Educational Research, 100*(4), 243–253. The study is ineligible for review because it does not examine the effectiveness of an intervention.

Bowen, E. W. (2006). *Accounting for agency in teaching mathematics: Understanding teachers' use of reform curriculum* (Unpublished master's thesis). Vanderbilt University, Nashville, TN. The study is ineligible for review because it does not include a student outcome.

Bush, W. S. (2005). Improving research on mathematics learning and teaching in rural contexts. *Journal of Research in Rural Education, 20*(8), 20–28. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Cai, J., Lew, H. C., Morris, A., Moyer, J. C., Fong Ng, S., & Schmittau, J. (2005). The development of students' algebraic thinking in earlier grades: A cross-cultural comparative perspective. *ZDM–The International Journal on Mathematics Education, 37*(1), 5–15. The study is ineligible for review because it does not include a student outcome.

Casey, B., Erkut, S., Ceder, I., & Young, J. M. (2007). Use of a storytelling context to improve girls' and boys' geometry skills in kindergarten. *Journal of Applied Developmental Psychology, 29*, 29–48. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Clements, D. H. (2007). Curriculum research: Toward a framework for "research-based curricula." *Journal for Research in Mathematics Education, 38*(1), 35–70. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Ebby, C. B. (2005). The powers and pitfalls of algorithmic knowledge: A case study. *Journal of Mathematical Behavior, 24*(1), 73–87. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Essex, N. K. (2006). Looking for gender differences in the mathematical work of elementary students. *Dissertation Abstracts International, 67*(12A) 204–4489. (UMI No. 3243791) The study is ineligible for review because it does not examine an intervention implemented in a way that falls within the scope of the review.

Feger, S., & Zibit, M. (2005). *The role of facilitation in online professional development: Engendering co-construction of knowledge.* Providence, RI: The Education Alliance at Brown University. The study is ineligible for review because it does not include a student outcome.

Fernandez, C., & Cannon, J. (2005). What Japanese and US teachers think about when constructing mathematics lessons: A preliminary investigation. *The Elementary School Journal, 105*(5), 481–498. The study is ineligible for review because it does not include a student outcome.

Flowers, J., Krebs, A. S., & Rubenstein, R. N. (2006). Problems to deepen teachers' mathematical understanding: Examples in multiplication. *Teaching Children Mathematics, 12*(9), 478. The study is ineligible for review because it does not include a student outcome.

Hands, L. (2006). Using classroom assessment to support growth of number sense in first grade. In S. Z. Smith, D. S. Mewborn, & M. E. Smith (Eds.), *Teachers engaged in research: Inquiry into mathematics classrooms, grades pre-K–2* (pp. 171–210). Greenwich, CT: Information Age Publishing. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Hill, H. C. (2005). Content across communities: Validating measures of elementary mathematics instruction. *Educational Policy, 19*(3), 447–475. The study is ineligible for review because it does not examine the effectiveness of an intervention.

Hundley, K. L. (2006). *Teacher efficacy in relation to mathematics education reform: An examination of a professional development study group of elementary teachers* (Unpublished master's thesis). Brigham Young University, Provo, UT. The study is ineligible for review because it does not include a student outcome.

Junk, D. L. (2006). *Teaching mathematics and the problems of practice: Understanding situations and teacher reasoning through teacher perspectives* (Unpublished doctoral dissertation). The University of Texas at Austin. The study is ineligible for review because it does not include a student outcome.

Kamina, P. (2006, November). *How fifth grade teachers used Investigations in Number, Data, and Space: A standards-based curriculum.* Paper presented at the 28th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education (PME-NA), Merida, Mexico. The study is ineligible for review because it does not include a student outcome.

Additional sources:

Kamina, P., & Tinto, P. (2005, October). *Lesson study: A case of the "Investigations" mathematics curriculum.* Paper presented at the 27th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education (PME-NA), Roanoke, VA.

Kamina, P. A. O. (2005). Teachers' perceptions and practices of inquiry-based instruction: A case study of fifth grade "Investigations" curriculum in an urban school. *Dissertation Abstracts International, 66*(05A), 229-1684.

Klein, D. (2007). School math books, nonsense, and the National Science Foundation. *American Journal of Physics, 75*, 101–102. The study is ineligible for review because it does not examine the effectiveness of an intervention.

Kniss, K. B. (2007). The effects of Investigations in Number, Data, and Space on the performance of at-risk students. *Masters Abstracts International, 46*(02), 68-605. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Lalli, C. B., & Feger, S. (2005). *Gauging and improving interactions in online seminars for mathematics coaches*. Providence, RI: The Education Alliance at Brown University. The study is ineligible for review because it does not include a student outcome.

Lehrer, R., & Schauble, L. (2005). Developing modeling and argument in the elementary grades. In T. A. Romberg, T. P. Carpenter, & F. Dremock (Eds.), *Understanding mathematics and science matters* (pp. 29–54). Mahwah, NJ: Lawrence Erlbaum Associates, Inc. The study is ineligible for review because it does not examine the effectiveness of an intervention.

Middleton, J. A., & Coleman, K. (2006). The development of leadership in mathematics: Cases of urban reform. In A. B. Danzig, K. M. Borman, B. A. Jones, & W. F. Wright (Eds.), *Learner-centered leadership: Research, policy, and practice* (pp. 131–148). Mahwah, NJ: Lawrence Erlbaum Associates, Inc. The study is ineligible for review because it does not examine the effectiveness of an intervention.

Noble, T., Nemirovsky, R., Wright, T., & Tierney, C. (2001). Experiencing change: The mathematics of change in multiple environments. *Journal for Research in Mathematics Education, 32*(1), 85–108. The study is ineligible for review because it does not use a comparison group design or single case design.

Mokros, J. (2003). Learning to reason numerically: The impact of Investigations. In S. L. Senk & D. R. Thompson (Eds.), *Standards-based school mathematics curricula: What are they? What do students learn?* (pp. 109 -131). Mahwah, NJ: Lawrence Erlbaum Associates, Inc. (This reference is for one of three separate studies included in the section: third- and fourth-grade students' number skills.) The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Reyes, W. G. (2007). *How integrating mathematics-based children's literature into the Investigations curriculum impacts students' acquisition of mathematical concepts and vocabulary in meaningful contexts* (Unpublished master's thesis). State University of New York College at Brockport. The study is ineligible for review because it does not examine the effectiveness of an intervention.

Rosebery, A. S., Warren, B., Ballenger, C., & Ogonowski, M. (2005). The generative potential of students' everyday knowledge in learning science. In T. A. Romberg, T. P. Carpenter, & F. Dremock (Eds.), *Understanding mathematics and science matters* (pp. 55–80). Mahwah, NJ: Lawrence Erlbaum Associates, Inc. The study is ineligible for review because it does not include an outcome within a domain specified in the protocol.

Ross, L. G. (2003). The effects of a standards-based mathematics curriculum on fourth and fifth grade achievement in two Midwest cities. *Dissertation Abstracts International, 64*(04), 1180A. (UMI No. 3088273) The study is ineligible for review because it does not use a comparison group design or single case design.

Schifter, D., Bastable, V., Russell, S. J., Seyferth, L., & Riddle, M. (2008). Algebra in the K–5 classroom: Learning opportunities for students and teachers. In C. E. Greenes & R. Rubenstein (Eds.), *Algebra and algebraic thinking in school mathematics: 70th yearbook* (pp. 263–267). Reston, VA: National Council of Teachers of Mathematics. The study is ineligible for review because it does not examine the effectiveness of an intervention.

Simpson, N. (2004). *Investigations in Number, Data, and Space evidence for success.* The study is ineligible for review because it does not use a comparison group design or a single-case design.

Slavin, R. E., & Lake, C. (2007). Effective programs in elementary mathematics: A best-evidence synthesis. *The Best Evidence Encyclopedia, 1*(2). Retrieved from http://www.bestevidence.org/word/elem_math_Feb_9_2007.pdf. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Smith, M. E. (2006). Introduction to the pre-K–2 volume. In S. Z. Smith, D. S. Mewborn, & M. E. Smith (Eds.), *Teachers engaged in research: Inquiry into mathematics classrooms, grades pre-K–2* (pp. 1–14). Greenwich, CT: Information Age Publishing. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Time, I. (2005). Math that matters. *Hands On, 27*(1), 1. The study is ineligible for review because it does not examine the effectiveness of an intervention.

Triantos, L. M. (2005). *The aftermath of implementing a standards-based curriculum in a K–8 district: Is there a correlation between hands-on instruction and math scores?* (Unpublished master's thesis). Rowan University, Glassboro, NJ. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Vaisenstein, A. (2006). A look at a child's understanding of mathematical ideas through his representations. In S. Z. Smith & M. E. Smith (Eds.), *Teachers engaged in research: Inquiry into mathematics classrooms, grades pre-K–2* (pp. 95–108). Greenwich, CT: Information Age Publishing. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Yelland, N. (2002). Creating microworlds for exploring mathematical understandings in the early years of school. *Journal of Educational Computing Research, 27*(1–2), 77–92. The study is ineligible for review because it does not take place in the geographic area specified in the protocol.

Yelland, N., & Masters, J. (2007). Rethinking scaffolding in the information age. *Computers & Education, 48*(3), 362–382. The study is ineligible for review because it does not take place in the geographic area specified in the protocol.

Appendix A.1: Research details for Agodini et al. (2010)

Agodini, R., Harris, B., Thomas, M., Murphy, R., & Gallagher, L. (2010). *Achievement effects of four early elementary school math curricula: Findings for first and second graders* (NCEE 2011-4001). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/pubs/20114001/pdf/20114001.pdf

Table A1. Summary of findings

Meets WWC evidence standards without reservations

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Mathematics achievement	110 schools/8,060 students	–1	No

Setting

The study took place in elementary schools in 12 districts across 10 states, including Connecticut, Florida, Kentucky, Minnesota, Mississippi, Missouri, Nevada, New York, South Carolina, and Texas. Of the 12 districts, three were in urban areas, five were in suburban areas, and four were in rural areas.

Study sample

Following district and school recruitment and collection of consent from all teachers in the participating grades, 111 participating schools were randomly assigned to one of four curricula: (a) *Investigations in Number, Data, and Space*®, (b) *Math Expressions*, (c) *Saxon Math*, and (d) *Scott Foresman–Addison Wesley Mathematics*. Blocked random assignment of the schools was conducted separately within each district. In each district, participating schools were grouped together into blocks of four to seven schools based on characteristics such as Title I eligibility, free or reduced-price lunch eligibility status, grade enrollment size, math proficiency, and proportion of White and Hispanic students. Two districts had an additional blocking variable (magnet school status in one district and year-round school schedule in another district). One district required that all schools that fed into the same middle school receive the same condition. Schools in each block were randomly assigned among the four curricula. On average, 11 students were randomly sampled from each participating classroom for assessment. One school with three teachers and 32 students assigned to *Math Expressions* withdrew from the study and did not permit posttesting of students at baseline or follow-up data collection.

The analysis sample included a total of 110 schools, 461 first-grade classrooms, 4,716 first graders, 328 second-grade classrooms, and 3,344 second graders. In the first grade sample, on average, 27 schools, 116 classrooms, and 1,180 students were assigned to each condition. In the second grade sample, on average, 18 schools, 82 classrooms, and 835 students were assigned to each condition.

Seventy-six percent of the schools in the study were eligible for Title I. Approximately half of the student population were eligible for free or reduced-price lunch. Across the schools, 39% of the student body were White, 32% were non-Hispanic Black, 26% were Hispanic, 2% were Asian, and 1% were American Indian or Alaskan Native.

Intervention group

Students used the *Investigations in Number, Data, and Space*® as their core math curriculum for one year. Study authors reported that 72% of the first-grade teachers and 80% of the second-grade teachers self-reported completing at least 80% of the curriculum.

Comparison group

There were three comparison curricula in the study: (a) *Math Expressions*, (b) *Saxon Math*, and (c) *Scott Foresman–Addison Wesley Mathematics*. Each curriculum was implemented by comparison teachers for one school year.

Math Expressions is published by Houghton Mifflin Harcourt and uses a blend of student-centered and teacher-directed instructional approaches. Students using the curriculum question and discuss mathematics and are explicitly taught effective procedures. There is an emphasis on using multiple specified objects, drawings, and language to represent concepts, and on learning through the use of real-world situations. Students are expected to explain and justify their solutions. Study authors reported that about nine out of 10 teachers self-reported completing at least 80% of the curriculum.

Saxon Math is published by Houghton Mifflin Harcourt and uses a teacher-directed approach that offers a script for teachers to follow in each lesson. It blends teacher-directed instruction of new material with daily practice of previously learned concepts and procedures. The teacher introduces concepts or efficient strategies for solving problems. Students receive instruction from the teacher, participate in guided practice, and then undertake individual practice. Frequent monitoring of student achievement is built into the program. Daily routines are extensive and emphasize practice of number concepts and use of methods (such as the use of number lines, counting on fingers, and diagrams) to represent mathematical concepts. Study authors reported that about six out of seven teachers self-reported completing at least 80% of the curriculum.

Scott Foresman–Addison Wesley Mathematics is published by Pearson Scott Foresman and is a curriculum that combines teacher-directed instruction with a variety of differentiated materials and instructional strategies. Teachers select the materials that seem most appropriate for their students. The curriculum is based on a consistent daily lesson structure, which includes direct instruction, hands-on exploration, the use of questioning, and practice of new skills. Study authors reported that about nine out of 10 teachers self-reported completing at least 80% of the curriculum.

Outcomes and measurement

Mathematics achievement was measured using the mathematics assessment developed for the ECLS-K class of 1998–99. The assessment is individually administered, nationally normed, and adaptive. According to the authors, the assessment meets accepted standards of validity and reliability. Scale scores from an item response theory (IRT) model were used in the analysis. The test was administered in the spring—that is, from one to six weeks before the end of the school year of program implementation. The test also was administered in the fall of the implementation year (that is, within four weeks of the first day of classes) to assess students' baseline math achievement. For a more detailed description of the outcome measure, see Appendix B.

Support for implementation

Teachers in all four groups were provided training by the curriculum publisher. Teachers assigned to *Investigations in Number, Data, and Space*® were provided one day of initial training in the summer before the school year began. Follow-up sessions were typically three to four hours long and held after school.

Teachers assigned to *Math Expressions* (comparison group 1) were provided two days of initial training in the summer before the school year began. Two follow-up trainings were offered during the school year. Follow-up sessions typically consisted of classroom observations followed by short feedback sessions with teachers.

Teachers assigned to *Saxon Math* (comparison group 2) were provided one day of initial training in the summer before the school year began. One follow-up training session, tailored to meet each district's needs, was offered during the school year.

Teachers assigned to *Scott Foresman–Addison Wesley Elementary Mathematics* (comparison group 3) received one day of initial training in the summer before the school year began. Follow-up training was offered about every four to six weeks throughout the school year. Follow-up sessions were typically three to four hours long and held after school.

Appendix A.2: Research details for Gatti & Giordano, 2010

Gatti, G., & Giordano, K. (2010). *Pearson Investigations in Number, Data, and Space efficacy study: Final report.* Pittsburgh, PA: Gatti Evaluation, Inc.

Table A2. Summary of findings — **Meets WWC evidence standards with reservations**

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Mathematics achievement	8 schools/333 students	+5	Yes

Setting The study was conducted in elementary and middle schools in four districts, one each in Arizona, Massachusetts, Oregon, and South Carolina. Among the four districts, two were in a small city, one was in a mid-sized city, and one was in a large city.

Study sample From the schools that responded to an invitation to participate in the study, researchers selected those with student populations that were ethnically and socioeconomically diverse and whose math achievement was comparable to average achievement for all schools in the state. The school sample included one elementary school and two middle schools (serving fifth grade) in Massachusetts, one elementary school in South Carolina, three elementary schools in Arizona, and one elementary school in Oregon.[6] In the first year of the study, the study schools randomly assigned first- and fourth-grade teachers and their students to either the *Investigations in Number, Data, and Space*® curriculum (intervention) or their regular curriculum (comparison). In the following study year, all schools tried to place students in the same condition to which they were randomly assigned in the first year of the study. However, teachers were allowed to choose their preferred curriculum in the second year under the condition that at least one teacher in each school implement the *Investigations in Number, Data, and Space*® curriculum and at least one teacher implement a comparison curriculum.

Outcome data were reported for only the 333 students who remained in their randomly assigned condition across both years and completed the baseline assessment. The analytic sample included 155 students from the first-grade cohort and 178 students from the fourth-grade cohort. Within the first-grade cohort, 99 students were assigned to the intervention group and 56 students were assigned to the comparison group. Among the fourth-grade cohort, 99 students were assigned to the intervention group and 79 students were assigned to the comparison group.

The size of the schools in the study ranged from 361 to 798 students. In all but one school, Caucasian and Hispanic/Native American students represented the two largest ethnic groups. Among all eight schools, Caucasians represented 34% to 76% of the student population, Hispanic/Native Americans represented 7% to 47%, African American/Caribbeans represented 2% to 37%, and Asian Americans represented 3% to 10%. The proportion of students receiving free or reduced-price lunch ranged between 39% and 77%.

Intervention group

Students in the intervention group received the *Investigations in Number, Data, and Space®* curriculum for two years. The teachers in the intervention group received a curriculum package that included the Implementation Guide, Resource Binder, Schools and Families Resource Book, Spanish Teaching Companion, Student Activity Book, Student Math Handbook, Manipulatives Kit, Cards Package, Access to Success.net with extra activities, Online Lesson Planner, Online Resource Masters, Online Student Handbook, and ExamView. On average, teachers delivered 65–69 minutes of the intended 70 minutes of daily instruction. All of the classrooms completed at least five of the intervention's nine thematic units.

Comparison group

Students in the comparison group received their existing curricula for two years. Two of the most widely-used math curricula, traditional skills-based programs, were taught to 86% of the comparison students. The other students received a math program created by their teachers from various sources. In one district, which had used the *Investigations in Number, Data, and Space®* curriculum in the past, comparison teachers also used selected manipulatives from this curriculum. In a second district, comparison teachers also supplemented their regular curriculum with selected activities and manipulatives from the *Investigations in Number, Data, and Space®* curriculum. Comparison teachers in the first- and second-grade classrooms delivered, on average, 56–58 minutes of daily math instruction. In the fourth- and fifth-grade classrooms, the comparison teachers delivered, on average, 60–68 minutes of daily math instruction.

Outcomes and measurement

The outcome for this study was the GMADE. The GMADE is a nationally normed, standardized test used to measure math achievement in grades K–12. Students were administered the GMADE in the first and last month of the school year in each of the two years. The outcome data at the end of Year 1 are presented in Appendix D. These findings do not contribute to the evidence rating because these Year 1 results do not represent the effect of the full (two-year) intervention as implemented by the study authors. For a more detailed description of the outcome measure, see Appendix B.

Support for implementation

The *Investigations* teachers participated in a day-long training offered by the publisher that focused on the key concepts of the curriculum and instructional practices. These teachers also participated in two to five additional meetings each year with curriculum experts, lasting from 30 minutes to four hours, to discuss upcoming units, state standards, special student needs, and program components.

In one district, the comparison teachers attended a half-day long training seminar. In a second district, teachers attended a half-day long workshop when the program was first adopted. Following adoption, the district math coordinator offered after school training sessions. Teachers were provided a pacing guide that correlated state standards with the textbook and were offered an additional 16 hours of teacher professional development by their schools. Teachers in a third district did not receive any training on the curriculum or professional development. In the fourth district, teachers attended a day-long training offered by the publisher when the curriculum was first adopted. The school offered additional sessions to the teachers which focused on addressing the weaker areas of the curriculum.

Appendix B: Outcome measures for mathematics achievement domain

Mathematics achievement	
Early Childhood Longitudinal Study–Kindergarten (ECLS-K) Math Assessment	This math assessment was developed for the ECLS-K class of 1998–99. The ECLS-K is a nationally normed adaptive test. The math assessment measures understanding and skills in five content areas: (a) number sense, properties, and operations; (b) measurement; (c) geometry and spatial sense; (d) data analysis, statistics, and probability; and (e) patterns, algebra, and functions. On the first-grade test, approximately three-quarters of the items focused on number sense, properties, and operations, with the remaining items predominantly drawn from the areas of data analysis, statistics, and probability; and patterns, algebra, and functions. An ECLS-K math assessment for the second grade did not exist, so the study authors worked with the developer of the ECLS-K, Educational Testing Service, to select appropriate items from existing ECLS-K math assessments (including the K–1, third-, and fifth-grade instruments). Half of the items in the second-grade test were related to number sense, properties, and operations, with the other half predominantly covering measurement; geometry and spatial sense; and patterns, algebra, and functions (as cited in Agodini et al., 2010).
Group Mathematics Assessment and Diagnostic Evaluation (GMADE)	This is a standardized, nationally normed, multiple choice test published by Pearson. The test includes three subtests: (a) concepts and communications (28 questions); (b) operations and computation (24 questions); and (c) process and applications (28 questions). There are nine levels of the GMADE assessments that span grades K–12, with two forms for each level. In this study, level 1 and level 4 tests were administered, with form A administered pre-intervention and form B administered post-intervention (as cited in Gatti & Giordano, 2010).

Appendix C: Findings included in the rating for the mathematics achievement domain

			Mean (standard deviation)		WWC calculations			
Outcome measure	**Study sample**	**Sample size**	**Intervention group**	**Comparison group**	**Mean difference**	**Effect size**	**Improvement index**	***p*-value**
Agodini et al., 2010[a]								
ECLS-K math	Grade 1 (versus *Math Expressions*)	54 schools/ 2,339 students	43.82 (8.04)	44.74 (8.52)	–0.92	–0.11	–4	0.01
ECLS-K math	Grade 1 (versus *Saxon Math*)	54 schools/ 2,235 students	44.69 (8.04)	45.23 (7.32)	–0.54	–0.07	–3	0.15
ECLS-K math	Grade 1 (versus *SFAW*)	57 schools/ 2,396 students	44.40 (8.04)	44.43 (8.15)	–0.03	0.00	0	0.93
ECLS-K math	Grade 2 (versus *Math Expressions*)	35 schools/ 1,638 students	70.84 (15.75)	71.38 (16.70)	–0.54	–0.03	–1	0.49
ECLS-K math	Grade 2 (versus *Saxon Math*)	36 schools/ 1,711 students	71.13 (15.75)	72.53 (16.16)	–1.40	–0.09	–3	0.09
ECLS-K math	Grade 2 (versus *SFAW*)	36 schools/ 1,623 students	71.66 (15.75)	70.31 (15.74)	1.35	0.09	+3	0.09
Domain average for mathematics achievement (Agodini et al., 2010)						**–0.04**	**–1**	**Not statistically significant**
Gatti & Giordano, 2010[b]								
GMADE	Grade 1 cohort at end of Year 2	6 schools/ 155 students	57.05 (10.21)	57.29 (12.81)	–0.24	–0.02	–1	0.89
GMADE	Grade 4 cohort at end of Year 2	8 schools/ 178 students	54.93 (13.26)	51.59 (13.28)	3.34	0.25	+10	< 0.01
Domain average for mathematics achievement (Gatti & Giordano, 2010)						**0.12**	**+5**	**Statistically significant**
Domain average for mathematics achievement across all studies						**0.04**	**+2**	**na**

Table Notes: For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the average change expected for all students who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average student's percentile rank that can be expected if the student is given the intervention. The WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of each study's domain average was determined by the WWC. na = not applicable. ECLS-K math = Early Childhood Longitudinal Study–Kindergarten math assessment. SFAW = Scott Foresman–Addison Wesley Mathematics. GMADE = Group Mathematics Assessment and Diagnostic Evaluation.

[a] For Agodini et al. (2010), a correction for multiple comparisons was needed and resulted in significance levels that differ from those reported in the original study. Specifically, the *p*-value of 0.01 was higher than the critical *p*-value; therefore, the WWC does not find the result statistically significant. The *p*-values presented here were reported in the original study. The authors used a different multiple comparison adjustment and also found that this result was not statistically significant when adjusted for multiple comparisons. The intervention group mean is the unadjusted comparison group mean plus the program coefficients from the hierarchical linear modeling (HLM) analysis. The standard deviations for the intervention group are unadjusted standard deviations. This study is characterized as having an indeterminate effect because none of the effects are statistically significant or large enough to be considered substantively important according to WWC criteria. For more information, please refer to the WWC Standards and Procedures Handbook, version 2.1, p. 96.

[b] For Gatti & Giordano (2010), a correction for multiple comparisons was needed but did not affect significance levels. The *p*-values presented here were reported in the original study. The group means are adjusted means that controlled for differences in pretest scores, student demographics, and classroom environment indicators. The adjusted means and unadjusted standard deviations for both grades and the *p*-value for grade 1 were provided to the WWC by the authors. This study is characterized as having a statistically significant positive effect because the effect for at least one measure within the domain is positive and statistically significant, and no effects are negative and statistically significant, accounting for multiple comparisons. For more information, please refer to the WWC Standards and Procedures Handbook, version 2.1, p. 96.

Appendix D: Description of supplemental findings from Year 1 for the mathematics achievement domain

			Mean (standard deviation)		WWC calculations			
Outcome measure	**Study sample**	**Sample size**	**Intervention group**	**Comparison group**	**Mean difference**	**Effect size**	**Improvement index**	***p*-value**
Gatti & Giordano, 2010[a]								
GMADE	Grade 1 at end of Year 1	6 schools/ 153 students	61.53 (9.19)	64.72 (11.72)	–3.19	–0.31	–12	0.03
GMADE	Grade 4 at end of Year 1	6 schools/ 177 students	51.57 (13.78)	55.55 (13.17)	–3.98	–0.29	–12	< 0.01

Table Notes: The supplemental findings presented in this table are additional Year 1 findings from the studies in this report. These findings do not factor into the determination of the intervention rating because these results do not represent the effect of the full (two-year) intervention as implemented by the study authors. For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the average expected for all students who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average student's percentile rank that can be expected if the student is given the intervention. GMADE = Group Mathematics Assessment and Diagnostic Evaluation.

[a] For Gatti & Giordano (2010), a correction for multiple comparisons was needed but did not affect significance levels. The *p*-values presented here were provided to the WWC by the authors. The group means are adjusted means for the Year 2 analysis sample, excluding three students, at the end of the first year of the study. The adjusted means controlled for differences in pretest scores, student demographics, and classroom environment indicators. The adjusted means and unadjusted standard deviations for both grades were provided to the WWC by the authors.

Endnotes

[1] The descriptive information for this program was obtained from a publicly-available source: the developer's website (http://investigations.terc.edu, downloaded March 2012). The WWC requests distributors review the program description sections for accuracy from their perspective. The program description was provided to the distributor in December 2011; however, the WWC received no response. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review. The literature search reflects documents publicly available by December 2011.

[2] The previous report was released in February 2009. This report has been updated to include reviews of six studies that have been released since 2009. Of the additional studies, two were not within the scope of the review protocol for the Elementary School Mathematics topic area, two were within the scope of the review protocol for the Elementary School Mathematics topic area but did not meet evidence standards, one meets evidence standards without reservations, and one meets evidence standards with reservations. A complete list and disposition of all studies reviewed are provided in the references. The studies in this report were reviewed using the Evidence Standards from the WWC Procedures and Standards Handbook (version 2.1), along with those described in the Elementary School Mathematics review protocol (version 2.0). The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

[3] Absence of conflict of interest: One of the studies summarized in this intervention report, Agodini et al. (2010), was prepared by staff of one of the WWC contractors. Because the principal investigator for the WWC review of Elementary School Mathematics is also a staff member of that contractor and a lead author of this study, the study was rated by staff members from a different organization. The report was then reviewed by the principal investigator, a WWC Quality Assurance reviewer, and an external peer reviewer.

[4] For criteria used in the determination of the rating of effectiveness and extent of evidence, see the WWC Rating Criteria on p. 19. These improvement index numbers show the average and range of student-level improvement indices for all findings across the studies.

[5] All cost information was obtained from the developer's website in December 2012.

[6] Grade, delivery method, and program type refer to the studies that meet WWC evidence standards without or with reservations.

[7] For the fourth-grade cohort, the study followed these students through the fifth grade. Two middle schools are included in the study because fifth grade in Massachusetts is offered in middle schools. In one middle school, two teachers used the *Investigations in Number, Data, and Space®* curriculum in all of their sections. In the second middle school, two teachers used their regular math curriculum.

[8] Consistent with the WWC practice of reporting significant results prior to insignificant results, the results for the fourth-grade cohort are presented first in this section of the report. In the remainder of the report, the descriptive information and results for the two cohorts are presented in grade order.

[9] The Gatti and Giordano (2010) study's findings for students after one year of program exposure are presented in Appendix D.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse (2013, February). *Elementary School Mathematics intervention report: Investigations in Number, Data, and Space®*. Retrieved from http://whatworks.ed.gov.

WWC Rating Criteria

Criteria used to determine the rating of a study

Study rating	Criteria
Meets WWC evidence standards without reservations	A study that provides strong evidence for an intervention's effectiveness, such as a well-implemented RCT.
Meets WWC evidence standards with reservations	A study that provides weaker evidence for an intervention's effectiveness, such as a QED or an RCT with high attrition that has established equivalence of the analytic samples.

Criteria used to determine the rating of effectiveness for an intervention

Rating of effectiveness	Criteria
Positive effects	Two or more studies show statistically significant positive effects, at least one of which met WWC evidence standards for a strong design, AND No studies show statistically significant or substantively important negative effects.
Potentially positive effects	At least one study shows a statistically significant or substantively important positive effect, AND No studies show a statistically significant or substantively important negative effect AND fewer or the same number of studies show indeterminate effects than show statistically significant or substantively important positive effects.
Mixed effects	At least one study shows a statistically significant or substantively important positive effect AND at least one study shows 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, OR At least one study shows a statistically significant or substantively important effect AND more studies show an indeterminate effect than show a statistically significant or substantively important effect.
Potentially negative effects	One study shows a statistically significant or substantively important negative effect and no studies show a statistically significant or substantively important positive effect, OR Two or more studies show statistically significant or substantively important negative effects, at least one study shows a statistically significant or substantively important positive effect, and more studies show statistically significant or substantively important negative effects than show statistically significant or substantively important positive effects.
Negative effects	Two or more studies show statistically significant negative effects, at least one of which met WWC evidence standards for a strong design, AND No studies show statistically significant or substantively important positive effects.
No discernible effects	None of the studies shows a statistically significant or substantively important effect, either positive or negative.

Criteria used to determine the extent of evidence for an intervention

Extent of evidence	Criteria
Medium to large	The domain includes more than one study, AND The domain includes more than one school, AND The domain findings are based on a total sample size of at least 350 students, OR, assuming 25 students in a class, a total of at least 14 classrooms across studies.
Small	The domain includes only one study, OR The domain includes only one school, OR The domain findings are based on a total sample size of fewer than 350 students, AND, assuming 25 students in a class, a total of fewer than 14 classrooms across studies.

Glossary of Terms

Attrition Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. The WWC considers the total attrition rate and the difference in attrition rates across groups within a study.

Clustering adjustment If intervention assignment is made at a cluster level and the analysis is conducted at the student level, the WWC will adjust the statistical significance to account for this mismatch, if necessary.

Confounding factor A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor.

Design The design of a study is the method by which intervention and comparison groups were assigned.

Domain A domain is a group of closely related outcomes.

Effect size The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes.

Eligibility A study is eligible for review and inclusion in this report if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.

Equivalence A demonstration that the analysis sample groups are similar on observed characteristics defined in the review area protocol.

Extent of evidence An indication of how much evidence supports the findings. The criteria for the extent of evidence levels are given in the WWC Rating Criteria on p. 19.

Improvement index Along a percentile distribution of students, the improvement index represents the gain or loss of the average student due to the intervention. As the average student starts at the 50th percentile, the measure ranges from –50 to +50.

Multiple comparison adjustment When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary.

Quasi-experimental design (QED) A quasi-experimental design (QED) is a research design in which subjects are assigned to intervention and comparison groups through a process that is not random.

Randomized controlled trial (RCT) A randomized controlled trial (RCT) is an experiment in which investigators randomly assign eligible participants into intervention and comparison groups.

Rating of effectiveness The WWC rates the effects of an intervention in each domain based on the quality of the research design and the magnitude, statistical significance, and consistency in findings. The criteria for the ratings of effectiveness are given in the WWC Rating Criteria on p. 19.

Single-case design A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention.

Standard deviation The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample tend to be spread out over a large range of values.

Statistical significance Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ($p < 0.05$).

Substantively important A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.

Please see the WWC Procedures and Standards Handbook (version 2.1) for additional details.