# Effective Programs in Elementary Mathematics:

# A Best-Evidence Synthesis

Marta Pellegrini

University of Florence

Cynthia Lake

Amanda Inns

Robert E. Slavin

Johns Hopkins University

October, 2018

Author Note: Marta Pellegrini, Department of Education and Psychology, University of Florence; Cynthia Lake, Center for Research and Reform in Education, Johns Hopkins University; Amanda Inns, Center for Research and Reform in Education, Johns Hopkins University; Robert E. Slavin, Center for Research and Reform in Education, Johns Hopkins University.

Correspondence concerning this article should be addressed to Robert E. Slavin, rslavin@jhu.edu.

*The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Research and Reform in Education (CRRE) under funding from the Institute of Education Sciences, U.S. Department of Education.*

**Abstract**

This article reviews research on the mathematics achievement outcomes of all programs with at least one study meeting inclusion criteria. 78 studies evaluated 61 programs in grades K-5. The studies were very high in quality, with 65 (83%) randomized and 13 (17%) quasi-experimental evaluations. Programs were organized in 8 categories. Particularly positive outcomes were found for tutoring programs. One-to-one and one-to-small group models had equal impacts, as did teachers and paraprofessionals as tutors. Technology programs showed modest positive impacts. Professional development approaches focused on helping teachers gain in understanding of math content and pedagogy had no impact on student achievement, but more promising outcomes were seen in studies focused on instructional processes, such as cooperative learning. Whole-school reform, social-emotional approaches, math curricula, and benchmark assessment programs found few positive effects, although there were one or more effective individual approaches in most categories. The findings suggest that programs emphasizing personalization, engagement, and motivation are most impactful in elementary mathematics instruction, while strategies focused on textbooks, professional development for math knowledge or pedagogy, and other strategies that do not substantially impact students' daily experiences have little impact.

Key words: Mathematics, elementary school, math programs, research review, best-evidence synthesis.

## Effective Programs in Elementary Mathematics: A Best-Evidence Synthesis

Whenever political and educational leaders speak or write about how the economic success of nations depends on the academic success of its students, the first example mentioned is often mathematics (e.g., American Institutes for Research [AIR], 2006; National Research Council, 2004). The U.S. is falling behind in the world, we are often told, because our Asian and European peers are surpassing us in mathematics, and will therefore soon surpass us in the ability to manage and profit from new technologies, increasingly complex science, and advanced engineering, all of which depend on a solid core of skill in mathematics.

If mathematics is essential to success in all quantitative endeavors and occupations, then success in elementary mathematics is of particular importance. In elementary school, students learn the basic mathematical ideas and operations, of course, but they also learn that they are either "good at math" or "not good at math". They also learn that math is fun and worthwhile, or that it is tedious and unrewarding. These early learnings can have long term consequences.

Students from disadvantaged homes, and African American, Hispanic, Native American students in particular, are likely to perform poorly in elementary math. Believing they are not good at math, many avoid math activities and, ultimately, professions that require mathematical proficiency (Barton & Coley, 2010; Ganley & Lubienski, 2016).

According to the 2017 National Assessment of Educational Progress (NAEP; National Center for Educational Statistics [NCES], 2018), mathematics scores for U.S. fourth graders increased steadily from 1990 to 2000 for all groups. However, from 2003 to 2017, scores have remained relatively flat, averaging 40% proficient in 2017. More distressingly, gaps in math performance have remained consistent all the way back to 1990. In 2017, 51% of White

students but only 19% of African Americans, 26% of Hispanics, and 24% of Native Americans scored at the proficient level. Asian-American students did very well, averaging 67% proficient in 2017. 57% of students not qualifying for free- or reduced-price lunch scored proficient on NAEP, while only 25% of students who qualified for free lunch did so. The gap between boys and girls was relatively small, but it has not diminished in recent years. Boys averaged 42% proficient and girls 38% in 2017.

In light of the importance of elementary mathematics and the social and economic implications of ethnicity, social class, and gender gaps, it is clear that substantial investments in improving elementary mathematics are justified. Yet which programs and practices are most likely to increase achievement and reduce gaps?

The importance of evidence for the effectiveness of mathematics programs has increased for U.S. schools as a consequence of the 2015 Every Student Succeeds Act (ESSA). ESSA defines three levels of evidence that have important consequences for certain federal funding, especially for low-achieving schools. All three require at least one significant positive effect and no significant negative effects in well-implemented studies. "Strong" requires at least one positive randomized study, "moderate" requires at least one positive quasi-experimental or matched study, and "promising" requires at least one positive correlational study, with controls for pretests or other covariates. ESSA policies have heightened policy interest in evidence for this reason. This review focuses on the mathematics outcomes evaluating elementary math programs that used research designs and measures most likely to qualify programs for the "strong" and "moderate" ESSA categories.

**Need for This Review**

In recent years, several reviews on elementary math programs have been published.

Slavin and Lake (2008) identified 87 rigorous studies of outcomes of elementary math programs and concluded that math programs that incorporate cooperative learning, classroom management and motivation, and tutoring had the most positive effects on math student achievement. Another review of experimental studies by Jacobse and Harskamp (2011) examined the impact of math interventions in grades K-6 and identified forty studies. The authors reported that small group or individual interventions had greater effects on math achievement than did whole-class programs. Dividing the interventions into those that use direct instruction and those that follow a constructivist approach of guiding students, the review showed no significant difference in outcomes.

Savelsbergh et al. (2016) carried out a meta-analysis of the effect of innovative science and math interventions on student achievement in grade 1 to 12. For mathematics, 19 studies were included. The authors distinguished four types of educational approaches: context-based teaching, inquiry-based learning, technology-rich learning environments, and collaborative learning. Interventions using technology found moderate positive effects.

Some recent reviews have focused on specific teaching methods in mathematics. Carbonneau, Marley, and Selig (2013) evaluated the efficacy of the use of manipulatives in elementary and secondary schools. The review identified 55 studies and found a small- to medium-sized effect on student achievement. Two other reviews focused on the efficacy of programs using technology in elementary and secondary schools. Li and Ma (2010) included 46 studies and found that computer technology had a greater effect when it was combined with a constructivist approach to teaching. Cheung and Slavin (2013) identified 45 rigorous studies of technology applications in mathematics carried out in elementary school settings and distinguished three different types: computer-managed learning, comprehensive models,

and supplemental computer-assisted instruction (CAI) technology. Supplemental CAI had the largest effect on mathematics achievement, with a mean effect size of +0.18.

Although several reviews of elementary mathematics programs have been carried out, the need for high-quality evaluations has particularly increased in recent years. The Institute for Education Sciences (IES), Investing in Innovation (i3) (recently supplanted by a similar program called Education Research and Innovation, or EIR), and England's Education Endowment Foundation (EEF), have funded numerous rigorous studies of elementary mathematics approaches. IES, i3/EIR, and EEF have insisted on randomized research designs, greatly expanding the number of studies using them. Other funders, including publishing and software companies, have also supported rigorous research evaluating elementary math programs. The great majority of studies, whether funded by government or by publishers, have used independent third-party evaluators. As noted earlier, the ESSA evidence standards have provided further incentives for developers and researchers to use rigorous designs. This review is focused on those studies that meet very high evidence standards, now available in large enough numbers to make such a review possible.

**Focus of the Review**

This review examines research on the effects of elementary math programs on student mathematics achievement. The purpose is to provide reliable information on the effectiveness of each program from rigorous experimental evaluations. The review considers the strength of evidence supporting particular programs, but it also groups interventions in categories based on their main components to find patterns that may have broader applicability. Analyses of categories and moderating factors (such as research designs and demographics) contribute to the advancement of theory as well as practical understanding.

## Method

The present review uses best evidence synthesis (Slavin, 1986), a form of meta-analysis (Glass, 1976; Lipsey & Wilson, 2001) that adds to systematic review procedures a narrative description of the context, design, and findings of each qualifying study.

### Inclusion Criteria

The review used rigorous inclusion criteria designed to minimize bias and provide educators and researchers with reliable information on programs' effectiveness. Inclusion criteria were as follows:

1. Studies had to evaluate student mathematics outcomes of programs for elementary schools, grades K-5. Sixth graders were also included if they were in elementary schools. Studies from England involved primary schools with students from Reception (U.S. kindergarten), and Years 1 to 6. Students who qualified for special education services but attended mainstream mathematics classes were included.
2. Studies had to use experimental methods with random assignment to treatment and control conditions, or quasi-experimental (matched) methods in which treatment assignments were specified in advance. Studies that matched a control group to the treatment group after posttest outcomes were known (post-hoc quasi-experiments or ex post facto designs) were not included.
3. Studies had to compare experimental groups using a mathematics program to control groups using an alternative program already in place, or "business-as-usual".
4. Studies of evaluated programs had to be delivered by ordinary teachers, not by the program developers, researchers, or their graduate students.
5. Studies had to provide pretest data. If the pretest differences between experimental and

control groups were greater than 25% of a standard deviation, the study was excluded. Pretest equivalence had to be acceptable both initially and based on pretests for the final sample, after attrition.

6. Studies with differential attrition between experimental and control groups from pre- to post-test of more than 15% were excluded.
7. Studies' dependent measures had to be quantitative measures of mathematics performance. Assessments made by developers of the program or researchers were excluded, as such measures have been found to overstate program impacts (Cheung & Slavin, 2016; de Boer, Donker, & van der Werf, 2014; Pellegrini, Inns, Lake, & Slavin, 2018).
8. Studies had to have at least two teachers and 30 students in each condition (Inns, Pellegrini, Lake, & Slavin, 2018b).
9. Studies had to have a minimum duration of 12 weeks, to make it more likely that effective programs could be replicated over extended periods.
10. Studies could have taken place in any country, but the report of the study had to be available in English. In practice, all studies took place in the U.S., U.K., Canada, or Germany.
11. Studies had to have been carried out after 1990, but for technology approaches we used a start date of 2000, due to the significant advances in technology since that date.

**Literature Search and Selection Procedures**

A broad literature search was carried out in an attempt to locate every study that might meet the inclusion requirements. Then studies were screened to determine whether they were eligible for review. The process is summarized in Figure 1. It used a multi-step

process that included (a) an electronic database search, (b) a hand search of key peer-reviewed journals, (c) an ancestral search of recent meta-analyses, (d) a web-based search of educational research sites and educational publishers' sites, and (e) a final review of citations found in relevant documents retrieved from the first search wave.

First, electronic searches were conducted in educational databases (JSTOR, ERIC, EBSCO, PsycINFO, ProQuest Dissertations & Theses Global) using different combinations of key words (e.g., "elementary students," "mathematics", "achievement", "effectiveness", "RCT", "QED"). Search results were limited to studies published between 1990 and March 2018, or between 2000 and March 2018 for technology approaches.

We also searched in recent tables of contents of seven key mathematics and general educational journals from 2013 to 2018: *American Educational Research Journal, Journal of Educational Psychology, Journal of Research on Educational Effectiveness, The Elementary School Journal, Journal for Research in Mathematics Education, Learning and Instruction,* and *Review of Educational Research.*

Following the search in educational journals, we investigated citations from previous reviews of elementary mathematics programs or related topics such as technology applications and tutoring (Cheung & Slavin, 2013; Jacobse & Harskamp, 2011; Li and Ma, 2010; Savelsbergh et al., 2016; Slavin & Lake, 2008).

In addition, we conducted searches by program name, examined the websites of educational publishers, and contacted producers and developers of mathematics programs to check for studies we had missed. We were particularly careful to be sure we found unpublished as well as published studies, because of the known consequences of

publication bias in research reviews (Chow & Ekholm, 2018; Rothstein, Sutton, & Borenstein, 2006). We searched for studies published online by funding agencies such as i3, IES, and EEF, and for studies reviewed by the What Works Clearinghouse (WWC) and Evidence for ESSA. We also visited the websites of educational societies (American Educational Research Association and Society for Research on Educational Effectiveness) to search for conference presentations. Finally, we reviewed citations of documents retrieved from the first wave to search for any other studies of interest.

A first screen of each study was carried out by examining the title and abstract using inclusion criteria. Studies that could not be eliminated in the screening phase were located and the full text was read by one of the study authors. We retained the studies that met the inclusion criteria and those where inclusion was possible but not clear. All the studies retained were examined by a second author to confirm that they met the inclusion criteria. When the two authors were in disagreement the inclusion or exclusion of the study was discussed with a third author until consensus was reached.

After removing duplicate studies, these search strategies yielded 9,144 studies for screening. The screening phase eliminated 8,452 studies, leaving 692 full-text articles to be assessed for eligibility. Of these full-text articles that were reviewed, 614 studies did not meet the inclusion criteria, leaving 78 studies included in this review.

**Coding**

Studies that met the inclusion criteria were coded by one of the authors of the review. Then codes were verified by another author. As for the inclusion of the studies, disagreements were discussed with a third author until consensus was reached.

Data coded included: program components, study design, study duration, sample

size, grade level, participant characteristics, outcome measures, and effect sizes.

We also identified variables that could possibly moderate the effects in the review. We coded moderators concerning methodological features, such as research design (e.g., randomized vs. quasi-experiment) and those concerning study characteristics, such as grade level (K-2 vs. 3-6), student achievement levels (low achievers vs. average/high achievers), socio-economic status (low SES vs. moderate/high SES). For tutoring programs we also coded the type of implementer (teacher vs. paraprofessional) and the group size (one-to-one or one-to-small group).

**Effect Size Calculations and Statistical Procedures**

Effect sizes were computed as the mean difference between the posttest scores for individual students in the experimental and control groups after adjustment for pretests and other covariates, divided by the unadjusted standard deviation of the control group's posttest scores. Procedures described by Lipsey and Wilson (2001) were used to estimate effect sizes when unadjusted standard deviations were not available. Studies often reported outcomes on more than one measure. Since these outcome measures were not independent, we produced an overall average effect size for each study.

Statistical significance is reported for each study using procedures from the What Works Clearinghouse (WWC, 2017). If assignment to the treatment and control groups was at the individual student level, statistical significance was determined by using analysis of covariance (ANCOVA), controlling for pretests and possibly other factors, or using equivalent procedures, such as multiple regression. If assignment to the treatment and control groups was at the cluster level (e.g., classes or schools), statistical significance was determined by using multilevel modeling such as Hierarchical Linear Modeling (HLM,

Raudenbush & Bryk, 2002). Studies with cluster assignments that did not use HLM but mistakenly used student-level analysis were re-analyzed to estimate the significance with a formula provided by the WWC (2017) to account for clusters. Because of this, studies reported in the past as statistically significant based on individual-level analyses in cluster designs are now no longer reported as statistically significant.

Mean effect sizes across studies were calculated after assigning each study a weight based on inverse variance (Lipsey & Wilson, 2001), with adjustments for clustered designs suggested by Hedges (2007). In combining across studies and in moderator analysis, we used random-effects models as recommended by Borenstein, Hedges, Higgins, and Rothstein (2009) when there is reason to believe that there is no single "true" effect size, but a range of effect sizes. Weighted mean effect sizes and meta-analytic tests (Q statistics) were calculated for each program and category in R (R Core Team, 2016) using the *metafor* package (Viechtbauer, 2010).

**Limitations**

This review is focused on rigorous experimental studies evaluating student mathematics outcomes. Although other research designs, such as qualitative and correlational research, can add depth and understanding of the effects of mathematics programs, for policy purposes it is crucial to evaluate programs according to impacts on quantitative measures in rigorous designs. This is especially important in the U.S. in light of the congressionally-mandated ESSA evidence standards. Further, the review focuses on studies that took place in real school settings over a period of at least 12 weeks, without considering more theoretically-driven brief studies that may also provide useful information to researchers. The purpose, therefore, is to provide evidence on replicable programs evaluated in authentic school settings. Such

evaluations are particularly useful to educators. Finally, the review excludes measures made by researchers or developers of the programs. These measures may be of theoretical interest, but are often unfair to control groups because they are aligned with the content taught in the experimental but not in the control group, and may greatly overstate program outcomes (Cheung & Slavin, 2016; de Boer et al., 2014; Pellegrini et al., 2018).

## Categories of Mathematics Programs

Studies that met the inclusion criteria were divided into categories according to the main and most distinctive components of the programs. Category assignments were based on independent reading of articles and websites by the authors. All authors read all accepted studies, and if there were disagreements about categorizations they were debated and determined by consensus among all authors.

**Research and theory supporting main program components**. The categorization of the programs was guided by two main sources. The first was a report of the National Council of Teachers of Mathematics (NCTM, 2014) that provided guiding principles for school mathematics. It highlighted the importance of the use of effective curricula, technologies, professional development for teachers, and student assessments. It also emphasized that it is essential to provide students with low achievement in math with strong support and consistent opportunities to learn. The second was a report by the American Institutes for Research (AIR, 2006). Although the AIR report did not focus mainly on elementary schools, it identified strategies useful for addressing challenges to math performance in all grade levels. The recommendations of these two reports were adapted to identify seven categories of programs. The categories and their theoretical rationales were as follows.

***1. Tutoring.*** Following a great deal of research showing positive outcomes of one-to-

one or one-to-small group tutoring in reading (e.g., Galuschka, Ise, Krick, & Schulte-Körne 2014; Inns, Lake, Pellegrini, & Slavin, 2018a; Slavin, Lake, Davis, & Madden, 2011; Wanzek et al., 2016), several programs using similar strategies have been devised in elementary math, especially in the early grades. Tutoring may involve one teacher or one paraprofessional (teaching assistant) with one student, or one teacher or paraprofessional with a very small group of students, usually from two to five at a time.

There are several ways in which tutoring is likely to improve student math outcomes. First, tutoring (especially one-to-one) permits tutors to completely adapt their instruction to the needs of the student(s). Well-trained tutors are able to start with struggling students where they are and move them forward rapidly, instead of leaving them to flounder in the regular class with challenges too far above their current capabilities.

Second, tutors are likely to be able to build close personal relationships with the tutored student(s), giving them attention and praise that many students crave. In small group tutoring, students may also build relationships with groupmates, which may allow for mutual assistance as well as motivation.

We found two approaches related to tutoring that were so different from ordinary one-to-one or one-to-small group tutoring that we treated them separately and did not average their outcomes with other approaches. One was a form of distance tutoring over the Internet, in which tutors from India or Sri Lanka tutored students in England. The second was cross-age peer tutoring, also studied in England, in which Year 5 students tutored Year 3 students. These are interesting, but distinct, and there was only one study of each.

***2. Programs Incorporating Technology*** use computers or other advanced technology to help teach students. Such programs are usually supplementary, so students receive both

teacher-led instruction and technology-focused instruction. We distinguished two subcategories within the technology category. One was CAI approaches, in which students are assessed, placed at their appropriate level, and then given exercises and ongoing assessments in a step-by-step sequence to move them forward as rapidly as possible (Cheung & Slavin, 2013). Examples include SuccessMaker and Accelerated Math. The rationale for these programs is that personalized instruction will give students just the content they need, without regard to what the rest of the class is doing. CAI emphasizes the benefits of providing instruction just above children's current level of functioning, within their zone of proximal development (Vygotsky, 1978). Other programs, such as Mathematics and Reasoning, ST Math, and Time to Know, use multimedia content to help students visualize mathematical ideas (Mayer, 2009) within a CAI context that provides material at students' instructional levels.

***3. Professional Development* for *Math Content and Pedagogy*** approaches provide intensive content-focused professional development (PD) and intend to advance teachers' understanding of current standards-based content and effective instructional strategies. The theory of action is that the best way to enhance learning is to give teachers knowledge about math content and about ways of explaining it, rather than new texts or new software alone (Ball & Cohen, 1999; Cohen & Hill, 2000; Desimone & Garet, 2015; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). In particular, there is growing consensus among math education experts that teachers' deep understanding of the content and of mathematics-specific teaching methods may lead to an improvement in classroom performance and student learning (Ball, Thames, & Phelps, 2008; Hill, Rowan, & Ball, 2005; Saxe & Gearhart, 2001).

***4. Instructional Process Programs*** provide teachers with professional development (and sometimes materials) to help them implement innovations in classroom organization and

management, such as cooperative learning, classwide behavior approaches, and individualized or personalized instruction. This category had the highest effect size in the Slavin & Lake (2008) review of elementary mathematics programs, but many of the studies in the earlier review were not included in the current review either because they were reported before 1990 or they did not meet inclusion standards that have been significantly toughened since that time.

***5. Whole-School Reform*** approaches also vary widely, but all focus on all grades and multiple subjects. The rationale is that schools can improve outcomes by getting all staff to work together to improve instruction. These programs focus on the organization and management of the entire school, rather than on the implementation of a number of specialized and isolated school improvement initiatives. Previous research on comprehensive school reforms has demonstrated that results are highly variable, but some approaches are effective when well implemented (see Borman, Hewes, Overman, & Brown, 2003).

***6. Social-Emotional Learning*** approaches are not solely focused on mathematics. These are whole-school reforms designed to enhance student motivation by helping them with interpersonal skills and improving student behavior. The rationale for these emphasizes the need to improve students' behavior, motivation, and ability to function successfully in the classroom to benefit their achievement in math as well as other subjects. Students with high social and emotional competencies are likely to succeed better in their school courses than their peers (Corcoran, Cheung, Kim, & Chen, 2017; Domitrovich, Durlak, Staley, & Weissberg, 2017; Farrington et al., 2012).

***7. Mathematics Curricula*** are essentially textbooks, which schools and districts often adopt in hopes of improving math content and improving alignment with state or national standards. Many textbooks also provide supplemental technology or other add-

ons.

***8. Benchmark Assessments*** consist of tests given periodically (three to five times a year) to find out how students are proceeding toward success on state standards. The rationale is to give teachers and school leaders early information on student performance so they can make changes well before state testing (Henderson, Petrosino, Guckenburg, & Hamilton 2007; Herman, Osmundson, & Dietel, 2010; Konstantopolous, Miller, van der Ploeg, & Li, 2016).

## Results

A total of 78 studies evaluating 61 programs met the inclusion standards of this review. When a report compared two different programs to a control group, we counted it as two studies. The studies included were of high methodological quality: 65 (83%) of the studies were randomized trials and 13 (17%) were quasi-experimental studies. Table 1 shows the weighted mean effect sizes for each category, and Tables 2 to 9 report the main characteristics and outcomes of the studies, grouping them by category.

### Tutoring Programs

Fifteen studies, summarized in Table 2, evaluated tutoring programs. Six of these evaluated face-to-face, one-to-one tutoring. An additional study evaluated one-to-one tutoring from tutors in India or Sri Lanka delivered online to students in the U.K., and another used cross-age peer tutoring. Seven studies evaluated programs taught to small groups. All of the programs except cross-age tutoring involved students with low achievement in mathematics and took place in the U.S. or the U.K. Overall, the weighted mean effect size for one-to-one face-to-face tutoring was +0.26 ($k$ = 6, $p$ <.001), while the one-to-one online tutoring program had an effect size of -0.03 and cross-age peer tutoring had an effect size of +0.02. One-to-one tutoring by certified teachers ($k$ = 2, $ES$ = +0.27) and by paraprofessionals ($k$ = 3, $ES$ = +0.23)

did not differ in outcomes. It is important to note that in general, paraprofessionals were relatively well qualified (e.g., most had bachelor's degrees). Also, both certified teachers and paraprofessionals used structured programs and received extensive professional development. One program used paid AmeriCorps volunteers as tutors, and the ES was +0.20 ($p < .01$). Tutoring to small groups had a mean effect size of +0.32 ($k = 7$, $p < .001$). Surprisingly, outcomes of one-to-one and one-to-small group tutoring using structured programs did not differ. One-to-small group programs that used certified teachers ($k$=1, $ES$=+0.34) did not differ in outcomes from one-to-small group approaches that used paraprofessionals as tutors ($k$=6, $ES$=+0.32). The number of studies in each category of tutoring was small, so these findings must be interpreted with caution, but it is interesting that while all forms of face-to-face tutoring by paid adults had quite positive impacts on achievement, the outcomes were nearly identical for one-to-one and one-to-small group approaches and for certified teachers and paraprofessionals as tutors. The programs and studies are described below.

**One-to-one tutoring provided by teachers**

***Math Recovery*** provides low achievers in mathematics with one-to-one tutoring. Four to five 30-minute sessions a week are delivered by trained teachers for approximately 12 weeks. A quasi-experimental study by Smith, Cobb, Farran, Cordray, and Munter (2013) found significant positive effects for 1st grade students ($ES$ = +0.24, $p < .001$).

***Numbers Count*** is a one-to-one tutoring program for low-performing students. It consists of daily 30-minute sessions delivered by trained teachers for a total of 12 weeks. A study of Numbers Count in England (Torgerson, Wiggins, Torgerson, Ainsworth, & Hewitt, 2013) randomly assigned 6-7 year old students to the experimental or the control group. The effect size was statistically significant ($ES$ = +0.33, $p < .001$).

**One-to-one tutoring provided by paraprofessionals**

***Catch Up® Numeracy*** is a one-to-one tutoring intervention for low-achieving students in mathematics. It is delivered in two 15-minute sessions per week for 30 weeks by trained teaching assistants. Sessions focus on different numeracy components such as counting verbally, counting objects, ordinal numbers, and word problems. A randomized study of Catch Up® Numeracy (Rutt, Easton, & Stacey, 2014) in Years 2 to 6 in England (equivalent to U.S. grades 1 to 5) found a significant positive effect (*ES* = +0.21, $p < .05$).

***Galaxy Math*** is designed to promote number knowledge for low achievers in math. One-to-one tutoring by paraprofessionals occurs 3 times per week in 30-minute sessions for 16 weeks. Fuchs et al. (2013) evaluated the program with 1st grade students using a student-level randomized control trial and found a significant mean effect size of +0.25 ($p < .01$).

***Pirate Math*** is a word-problem tutoring program delivered 3 times per week over 16 weeks by paraprofessionals in 20-30 minute sessions. Fuchs et al. (2010) evaluated the program in a randomized trial with 3rd graders. The mean effect size was statistically significant (*ES* = +0.37, $p < .05$).

**One-to-one tutoring provided by paid volunteers**

***MathCorps*** is an AmeriCorps program serving students at risk for math problems. It is delivered 60 minutes per week by trained volunteers who receive regular stipends as well as other benefits if they complete their one-year commitment satisfactorily. ServeMinnesota (2017) evaluated the program with 4th-6th graders using a randomized trial and found a significant effect size of +0.20 ($p < .01$). It is important to note that in the Inns et al. (2018) review of research on programs for struggling readers, paid volunteer tutors (such as AmeriCorps members) also produced very good student outcomes, but unpaid

volunteers had much lower effect sizes.

**One-to-one tutoring delivered online**

***Affordable Online Maths Tuition*** is a tutoring program in which students receive one-to-one math tutoring over the Internet from trained tutors in India and Sri Lanka. Students receive tutoring in 45-minute sessions per week over 27 weeks. A study in England evaluated this program with Year 6 students in a cluster-randomized trial and found no significant effects (*ES* = -0.03, *ns*).

**Cross-age tutoring**

***Shared Maths*** is an English cross-age tutoring approach in which older students (Year 5 or 6) work with younger students (Year 3 or 4) to find solutions to math problems. A cluster randomized trial of Shared Maths in England (Lloyd et al., 2015) found no significant effects on mathematics achievement (*ES* = +0.02, *ns*).

**One-to-small group tutoring provided by teachers**

***Number Rockets*** is designed to address the mathematics difficulties of students at risk. Teachers tutor groups of two or three students for approximately 40 minutes each lesson, focusing on concepts and operations involving whole numbers. Students receive approximately 48 lessons over six months. A cluster randomized trial (Gersten et al., 2015) found significant positive effects for 1st graders (*ES* = +0.34, $p < .001$).

**Small-group tutoring provided by paraprofessionals**

***FocusMATH*** is a small-group tutoring program for students who perform below grade level in mathematics. Paraprofessionals provide explicit instruction in math content, concepts, and procedures. In a student-level randomized trial the program was delivered to 3rd and 5th graders in 30-minute sessions 2 to 4 days a week for a year (Styers & Baird-Wilkerson,

2011). The effect size was statistically significant (*ES* = +0.24, *p* < .001).

***Fraction Face-Off!*** is a tutoring program primarily focused on the interpretation of fractions. Paraprofessionals provide lessons to pairs of 4th grade students 3 times a week for 35 minutes. Across two 12-week randomized studies, the weighted mean effect size was +0.51. Both studies found significant positive effects (Fuchs et al., 2016a, *ES* = +0.39, *p* < .005; Fuchs, Sterba, Fuchs, & Malone, 2016b, *ES* = +0.64, *p* < .005).

***Fusion Math*** focuses on whole-number concepts and skills. The program is delivered by paraprofessionals in small groups of approximately five students in 30- minute sessions three times per week over a period of 19 weeks. A randomized trial (Clarke et al., 2014) found positive but not significant effects for 1st graders (*ES* = +0.11, *ns*).

***ROOTS*** is a kindergarten program for low-performing students delivered in 20-minute small-group tutoring sessions 5 times a week for 4-5 months by paraprofessionals. Across two randomized trials in kindergartens the weighted mean effect size was +0.24. Doabler et al. (2016) found an effect size of +0.32 (p < .01); Clarke et al. (2016) found a significant effect size of +0.32 on TEMA-3 and no significant effects on two other measures. Both studies had non-significant negative outcomes on follow-up in first grade, however (*ES* = -0.12; *ES* = -0.20, respectively).

**Programs Incorporating Technology**

Fourteen studies evaluated programs that strongly emphasize use of technology (Table 3). Some of these programs use typical CAI strategies, such as SuccessMaker and Accelerated Math, as supplements to classroom instruction. These approaches assess students' strengths and weaknesses and then assign learning activities and exercises designed to fill the gaps, with regular assessments and feedback to students and teachers. Other programs, such as ST Math

and Time to Know, emphasize the use of visual media. Most technology approaches rotate students through technology and non-technology activities. Combining all studies of programs incorporating technology, the weighted mean effect size was +0.07 ($k$ = 14, $p$ = .05).

***Accelerated Math*** is a supplementary approach to mathematics instruction that uses computers to assess children's levels of performance, and then generates individualized assignments appropriate to their needs. The program focuses on foundational skills, especially computations. Across one cluster quasi-experiment and two cluster randomized trials (Lambert, Algozzine, & McGee, 2014; Lehmann & Seeber, 2005; Ysseldyke & Bolt, 2007), the mean effect size was +0.03. No study found significant effects.

***DreamBox Learning*** is a supplemental online math program that focuses on teaching numbers and operations, place value, and number sense. It provides feedback to teachers on student program use and progress. A within-school randomized trial by Wang and Woodworth (2011a) found a non-significant mean effect size of +0.11 for kindergarten and 1st grade students (*ns*), although a geometry measure found statistically significant positive effects.

***Educational Program for Gifted Youth (EGPY)*** is a computer-based instructional program that uses multimedia lessons to introduce math concepts and exercises to practice them. It also gives tutorial support to struggling students. A within-school randomized trial by Suppes, Holland, Hu, and Vu (2013) with 2nd to 5th graders found no significant effects (*ES* = -0.01, *ns*).

***Odyssey® Math*** is comprehensive math instructional software consisting of a web-accessed series of learning activities, assessments, and math tools. The software is intended to be used as the main curriculum or as a supplemental program. Wijekumar, Hitchcock, Turner, Lei, and Peck (2009) evaluated the program as supplemental instruction with fourth graders in

a cluster randomized trial and found no significant effects (*ES* = +0.02, *ns*).

***SuccessMaker*** is an integrated learning system that incorporates curriculum, management, and assessment in one package. The program provides supplemental instruction to students who work through exercises at their level while the system adjusts future lessons based on their performance. Across one cluster quasi-experiment, one cluster randomized trial, and one within-school randomized trial (Gatti, 2009; Gatti, 2013; Gatti & Petrochenkov, 2010), the weighted mean effect size was +0.24. In the within-school trial by Gatti (2013), the outcomes were significantly positive (*ES* = +0.33, $p < .05$), but the findings in the other studies were not significant.

***Waterford Early Learning*** is a computer-based program designed to reinforce classroom instruction. The program provides materials such as computer software, digital books, and a wide range of learning activities. A cluster randomized trial evaluated the use of Waterford Early Learning for two years in K to 2nd grades (Magnolia Consulting, 2012). The effect size was not significant (*ES* = +0.04, *ns*).

***Mathematics and Reasoning*** aims to develop students' understanding of the logical principles of mathematics. Teachers are trained for one day and lessons are all delivered through electronic resources, including PowerPoints and online games that the children can access at school and at home. Worth, Sizmur, Ager, and Styles (2015) evaluated the program with Year 2 students in a cluster randomized trial and found significant positive effects (*ES* = +0.20, $p = .03$).

***Reasoning Mind*** is an adaptive learning environment that emphasizes in-depth understanding of arithmetic and the early introduction of algebraic concepts. The program provides training and support for teachers. Wang and Woodworth (2011b) evaluated the

program in a within-school randomized trial in 2nd to 5th grades and found no significant effects (*ES* = -0.01, *ns*).

***ST Math*** is a supplemental online approach designed to teach math reasoning through spatial and temporal representations. The program consists of games that engage students in solving math problems 60-90 minutes each week. A cluster randomized trial of ST Math in 3rd to 5th grades (Rutherford et al., 2014) found no significant effects after 1 or 2 years (*ES* = +0.08, *ns*).

***Time to Know*** is a blended approach in which students use one-to-one laptops with interactive learning activities. The program provides teachers with ongoing professional development. A cluster quasi-experimental study by Rosen and Beck-Hill (2012) with 4th and 5th graders found a non-significant effect size of +0.31 (*ns*) at the cluster level.

**Professional Development for Math Content and Pedagogy**

Twelve studies evaluated twelve programs focused on teacher professional development to improve teachers' knowledge of math content and pedagogy (Table 6). The programs use different types of support for teachers such as workshops, training, continuous professional development, in-school support, and coaching. They may focus on improving teachers' content knowledge, content-specific pedagogy, general pedagogy, or some combination of these. Three of these approaches (Math Pathways and Pitfalls, Mathematics Mastery, and AMSTI) also provide classroom materials. The weighted mean effect size was +0.04 (*k* = 12, *ns*) for all professional development programs focused on math content and pedagogy, and no program reported significantly positive outcomes.

***Intel Math*** provides teachers with an 80-hour summer workshop focused on

enhancing teachers' understanding of grades K-8 mathematics topics, plus mathematics learning communities in which teachers review student work, and opportunities to review videos of each other's teaching. The total PD requires 93 hours. In a large cluster randomized trial by Garet et al. (2016), the mean effect size was significantly negative (*ES* = -0.06, $p < .05$).

***Classroom Assessment for Student Learning (CASL)*** is a professional development program implemented via teacher learning teams, in which teachers regularly discuss and reflect on program content and their experiences applying it in the classroom. CASL predominantly focuses on formative assessment. A cluster randomized trial by Randel et al. (2016) in fourth and fifth grades found no significant effects (*ES* = +0.01, *ns*).

***Using Data*** provides professional development and support to help teachers identify and solve student learning problems using achievement data. It can be delivered online or face-to- face. A cluster randomized study with fourth and fifth graders (Cavalluzzo et al., 2014) found no significant effects (*ES* = +0.01, *ns*).

***Math Solutions,*** based on the ideas of Marilyn Burns, provides professional development to teachers to help them learn math content, understand how children learn math, use formative assessment, and use classroom strategies to enable student problem solving. A cluster randomized study involving 105 teachers in 19 mostly African-American schools evaluated Math Solutions with fourth and fifth graders (Jacob, Hill, & Corey, 2017). Teachers in Math Solutions received four-day summer institutes each summer over the 3-year experiment. There were significant effects on some measures of mathematics knowledge for teaching, but no differences on state math tests (*ES* = +0.06, *ns).*

***Cognitively Guided Instruction (CGI)*** focuses on providing teachers with extensive

professional development to build on intuitive mathematics understandings they already have. In a cluster randomized study by Schoen, LaVenia, & Tazar (2018), CGI teachers of grades 1 and 2 received four days of workshops each summer and two 2-day followup sessions each school year, focusing on whole number operations, equality, and problem solving. A two-year study in 22 schools in Florida found no program effects in either grade (*ES* = 0.01, *ns*). A 1989 study by Carpenter et al. was very influential in its time, and had promising effects (*ES*=+0.24) that were, however, not statistically significant at the cluster level (and did not fall within this review's time limits).

***Alabama Math, Science, and Technology Initiative (AMSTI)*** focuses on professional development and in-school support for teachers. Three components of the program foster the use of inquiry-based instruction: a 10-day summer professional development session and training during the school year; access to program materials, manipulatives and technology; and mentoring for instruction by AMSTI specialists. A cluster randomized trial by Newman et al. (2012) in 4th and 5th grades found no significant effects (*ES* = +0.05, *ns*).

***EarlyMath*** is a professional development program designed to promote teacher attitudes, practice, and knowledge through a multi-year process. The program uses learning labs focused on math content knowledge and individual coaching to provide individual support. Using grade-level groups and cross-grade grouping, teachers study the state benchmarks and performance descriptors to integrate them into their math teaching. Reid, Chen, and McCray (2014) evaluated EarlyMath in a cluster quasi-experimental study with K to 2nd graders and found no significant outcomes (*ES* = +0.01, *ns*).

***Math Pathways and Pitfalls*** is a supplementary curriculum for K-8 students with a

particular focus on professional development. The program also provides lesson structure and materials focused on developing number concepts and operations. A cluster randomized trial (Heller, 2010) in 4th and 5th grades found no significant effects (*ES* = +0.06, *ns*).

***Mathematics Mastery*** is a whole-school approach that supports teachers with summer training, continuous professional development resources, and peer collaboration. A cluster randomized trial of Mathematics Mastery in England (Vignoles, Jerrim, & Cowan, 2015) with Year 1 students found no significant effects (*ES* = +0.10, *ns*).

***PBS TeacherLine*** is an online professional development approach designed to provide high-quality support and resources for K-12 teachers. Teachers also interact with peers in a virtual learning environment. Dominguez, Nicholls, and Storandt (2006) evaluated the program in a cluster randomized trial with 3rd to 5th graders and found an effect size close to zero (*ES* = +0.03, *ns*).

***Philosophy for Children*** engages students in group dialogues on philosophical issues to help them become more willing and able to question, reason, construct arguments, and collaborate with others. Teachers receive two days of training and in-school support for delivering the program. A cluster randomized trial conducted by Gorard, Siddiqui, and See (2015) evaluated the program in England with Year 5 students and found no significant effects (*ES* = +0.10, *ns*).

***Primarily Math*** is a math professional development program consisting of mathematics content and pedagogy courses taken over 13 months. The program aims to change teachers' beliefs and attitudes towards teaching and improve student math achievement. A cluster quasi-experiment by Kutaka et al. (2017) in K to 2nd grades found no

significant effects (*ES* = +0.14, *ns*).

**Instructional Process Programs**

Instructional process programs are professional development approaches focused on helping teachers use models such as cooperative learning, personalization, classroom management, and reflection. The difference between the previous category and this one is that professional development for math content and pedagogy is targeted primarily on improving teachers' knowledge and pedagogy, while instructional process programs intend to improve teachers' skills in classroom organization and management.

In the Slavin & Lake (2008) review, instructional process programs were the most effective approaches (in comparison to technology and math curricula). With the exception of tutoring, this is still the case. Across five studies of four diverse programs, the average effect size was +0.25 ($k = 5$, $p < .01$).

***Team Assisted Individualization (TAI)*** incorporates cooperative learning and individualized instruction to teach math in grades 2-6. Students progress as rapidly as they are able through individualized materials, working in partnership with mixed-ability teams to meet team goals. TAI provides teachers support using peer coaching and teacher collaboration in instructional planning. Across two cluster quasi-experimental studies of TAI, the weighted mean effect size was +0.03. One study by Stevens and Slavin (1995) found effects that were positive but not significant at the cluster level (*ES* = +0.20, *ns*), and a small quasi-experiment by Karper and Melnick (1993) found non-significant negative effects (*ES* = -0.09, *ns*).

There were also high-quality studies of TAI in the 1980s. These fell before the 1990 start date of this review, so they are not shown in the tables and are not averaged

with other findings, but they are important for a full understanding of TAI. Two cluster randomized studies of TAI were reported by Slavin & Karweit (1985). One found a mean effect size of +0.38 ($p < .05$) and the other, a mean of +0.28. Both were significant at the cluster level. A cluster quasi-experimental study by Slavin, Madden, & Leavey (1984) had a mean effect size of +0.19, also significant at the cluster level. In all three studies, outcomes were much larger on standardized tests of computations than on tests of concepts and applications.

***PAX Good Behavior Game*** is a behavior management approach in which students are divided into two or more teams. Teachers monitor student behavior, and when a member of a team violates class rules, the student's team receives a check mark on the whiteboard. Teams that have the fewest rule violations are declared winners of the game. Weis, Osborne, and Dean (2015) conducted a cluster quasi-experiment in 1st and 2nd grades to evaluate the program and found statistically significant positive effects on student math achievement (*ES* = +0.32, *p* = .03).

***Individualizing Student Instruction (ISI)*** is a program designed to provide students with math instruction at their level. Students are placed in about four homogeneous groups and teachers provide instruction targeted to the students' assessed needs. A study by Connor et al. (2018) in which classes were randomly assigned to treatments within schools evaluated ISI with second graders. Using a difference-in-differences analysis, students in ISI averaged +0.12 (*ns*) across two measures of math achievement.

***ReflectED*** is a program designed to improve students' metacognition. Students reflect individually each week on their learning and record their reflections on a tablet or

laptop. A cluster randomized trial by Motteram, Choudry, Kalambouka, Hutcheson, and Barton (2016) evaluated ReflectED and found non-significant positive outcomes at the cluster level (*ES* = +0.30, *ns*).

**Whole-School Reform Programs**

Whole-school reform interventions provide professional development to principals or leadership teams and teachers to advance student learning. The four programs, summarized in Table 8, are quite diverse. Some focus on leadership, others teacher bonuses based on their performance, and one on the use of proven programs. CDDRE, the program focusing on use of proven programs, obtained significant positive effects at the district level on math (*ES* = +0.15, $p < .05$), but the weighted mean effect size for all four programs was -0.01 (*k* = 4, *ns*).

***Center for Data-Driven Reform in Education (CDDRE)*** provides consultation with district and school leaders on strategic use of data and selection of programs with good evidence of effectiveness. The program helps schools explore all sources of data collected by the district and use them to identify key areas of weakness and then select proven programs targeted to their identified areas of need. A cluster randomized trial by Slavin et al. (2013) found positive effects for fifth graders at the district level after four years of the intervention (*ES* = +0.15, $p < .05$).

***McREL Balanced Leadership*** is based on the results of meta-analyses conducted by Waters, Marzano and McNulty (2003) that found a relationship between school leadership and student achievement. The leadership responsibilities (e.g. monitoring instruction, involvement in curriculum) identified by Marzano et al. are the key content of the program. A 2-year cluster randomized trial (Jacob, Goddard, Kim, Miller, & Goddard, 2015) with 3rd to 5th graders found no significant effects (*ES* = +0.03, *ns*).

***Success in Sight*** is a whole-school intervention designed to address each school's specific needs and engage leadership teams and teachers in school improvement practices. The program consists of large-group professional development, onsite mentoring with leadership teams, and distance learning and support. A cluster randomized trial (Wilkerson, Shannon, Styers, & Grant, 2012) found significantly negative effects after a 2-year intervention with 3rd to 5th graders (*ES* = -0.11, *p* = .02).

***Teacher Advancement Program (TAP)*** is a schoolwide intervention in which teachers can earn extra pay based on a combination of their contribution to student achievement and observed performance in the classroom. The program also provides weekly meetings of teachers and mentors and regular observations by a school leadership team to help teachers meet their performance goals. A cluster randomized trial by Glazerman and Seifullah (2012) found non-significant negative outcomes with 4th to 6th graders (*ES* = -0.06, *ns*).

**Social-Emotional Interventions**

Eight studies evaluated mathematics achievement outcomes of five social-emotional learning programs (Table 4). These had a weighted mean effect size of +0.03 (*k* = 8, *ns*).

***INSIGHTS*** is a comprehensive intervention in which teachers and parents work together to support students' ability to self-regulate. Teachers and parents attend several workshops over time to develop skills in supporting students' social-emotional development and self-regulation. O'Connor, Cappella, McCormick, and McClowry (2014) evaluated the program in a cluster randomized trial with K to 1st graders and found no significant effects (*ES* = +0.04, *ns*).

***Positive Action*** is a whole-school reform strategy designed to improve social-

emotional and achievement outcomes by building school climate, self-control, goal-setting, problem-solving, persistence, and other skills. The program consists of 140 15- to 20-minute lessons taught 4 days per week. Across two cluster randomized trials of Positive Action, the weighted mean effect size was +0.16. The outcomes in a study by Snyder et al. (2010) were significantly positive on the Hawaii Content and Performance Standard (*ES* = +0.22, *p* = .04), but outcomes in a Chicago study by Bavarian et al. (2013) were non-significant (*ES* = +0.17, *ns*).

***PATHS (Promoting Alternative Thinking Strategies)*** is a school-based social and emotional learning program for helping students to manage their behavior, understand their emotions, and work with others. It consists of lessons covering topics such as identifying feelings, controlling impulses, and understanding other people's perspectives. The program was evaluated in a 2-year cluster randomized trial in England with Year 5 and 6 students by the Manchester Institute of Education (MIE, 2015). The study found no significant effects (*ES* = 0.00, *ns*).

***Responsive Classroom*** focuses on enhancing social-emotional skills and academic learning in elementary students. The program is based on daily meetings to create a sense of classroom community and rules established by students to prevent problems. Across one cluster quasi-experiment and one cluster randomized trial of Responsive Classroom the weighted mean effect size was -0.06 (Rimm-Kaufman et al., 2007, 2014).

***Social Skills Improvement System-Classwide Intervention Program (SSIS-CP)*** is designed to improve prosocial behavior using strategies such as reinforcement, modeling, role-playing, and problem-solving. SSIS focuses on promoting specific social skills related to academic success. Across two cluster randomized trials evaluating SSIS, the weighted mean

effect size was +0.01 (DiPerna, Lei, Bellinger, & Cheng 2016; DiPerna, Lei, Cheng, Hart, & Bellinger, 2018).

**Mathematics Curricula**

Sixteen studies evaluated ten mathematics curricula, primarily textbooks (Table 5). Most of them had a duration of one or two years and involved large samples ($n > 250$). Across all qualifying studies, the weighted mean effect size was +0.06 ($k = 16$, $p = .07$), and only two of the programs had significant positive effects.

***Early Learning in Mathematics*** is s a core kindergarten mathematics program that provides 120 45-minute lessons in addition to daily 15-minute calendar activities. The program focuses on number operations, geometry, measurement, and vocabulary. A cluster randomized trial by Clarke et al. (2015) in kindergarten found non-significant positive effects (ES = +0.11, *ns*).

***enVisionMATH*** focuses on interactive learning and problem-based activities and uses frequent student assessments. It can be used in print or technology versions. Across four cluster randomized studies (Resendez & Azin, 2006; Resendez, Azin, & Strobel 2009; Resendez & Manley, 2005; Strobel, Resendez, & DuBose, 2017), the mean effect size was -0.02.

***Everyday Mathematics*** is a core curriculum for students in prekindergarten through grade 6. It emphasizes real-life problem solving, manipulatives, concept development, and use of technology. A 2-year cluster randomized trial by Vaden-Kiernan et al. (2015) in K to 5$^{th}$ grades found no significant effects (ES = -0.01, *ns*).

***GO Math!*** is a K-8 curriculum that provides teacher guides, student books, and digital resources. The program involves scaffolding to support students' metacognition, the use of graphic organizers, writing that helps students in processing and connecting new information,

and vocabulary to communicate mathematically. Eddy, Hankel, Hunt, Goldman, and Murphy (2014) evaluated the program in a cluster randomized control trial with 1st to 3rd graders and found no significant effects (ES = +0.01, ns).

***Investigations in Number, Data, and Space*** is a math curriculum that uses a student-centered approach encouraging metacognitive reasoning. The program provides thematic units of three to eight weeks in which students first investigate and then discuss problems and strategies. Across two cluster randomized trials (Agodini, Harris, Thomas, Murphy, & Gallagher, 2010; Gatti & Giordano, 2008), the weighted mean effect size was -0.07. Gatti and Giordano (2008) found a significantly negative effect (*ES* = -0.23, $p < .05$).

***JUMP Math*** is a highly-scaffolded, direct instruction approach that covers all strands of K to 8 math. The program is focused on problem-based learning. Teachers implement lessons contained in detailed, explicit teachers' guides to help move students towards discovering target concepts. A cluster randomized trial by Solomon et al. (2011) in 5th grade found positive but not significant results at the cluster level (*ES* = +0.23, *ns*).

***Math Connects*** is a mathematics curriculum that provides print and online content and resources. The program includes diagnostic, practice, and benchmark assessments. A cluster quasi- experimental study (Jordan, 2009) in 2nd and 4th grades found no significant effects on math achievement (*ES* = +0.02, *ns*).

***Math Expressions*** is a curriculum that uses a combination of teacher-directed and student-centered instructional activities. Each day begins with a set of routines such as calendar, money, and counting. Later, the teacher provides instruction to the whole class, and students then practice the new skills or concepts in pairs, small groups, or individually. The program

was evaluated in a large cluster randomized trial with 1st and 2nd graders (Agodini et al., 2010). The effects were significantly positive (*ES* = +0.11, $p < .05$).

***Math in Focus*** is an adaptation of an approach used in Singapore called My Pals Are Here! Maths (MPHM). The purpose of this curriculum is to achieve mastery of mathematics concepts, computational skills, and problem solving skills using a concrete-to- pictorial-to-abstract progression for each skill, in carefully scaffolded lessons. Across two cluster quasi-experiments and one cluster randomized trial, the weighted mean effect size was +0.24 (Educational Research Institute of America [ERIA], 2010; 2013; Jaciw et al., 2016).

***Saxon Math*** is a curriculum for kindergarten to fourth grade based on a teacher-directed instructional approach with scripted lesson plans. It is organized in five daily activities: morning routines, fact practice, an explicit lesson, guided class practice, and homework. The program was evaluated in a large cluster randomized trial with 1st and 2nd graders (Agodini et al., 2010). The effects were positive but not significant (*ES* = +0.11, *ns*).

**Benchmark Assessments**

Four studies evaluated programs that use benchmark assessments (Table 9). Overall, the weighted mean effect size was 0.00 (*k* = 4, *n.s.*).

***Achievement Network (ANet)*** focuses on use of academic assessments to improve teaching and learning. It includes quarterly interim assessments in English and math, data tools including reports on students' progress, coaching of school leaders to support their teachers' use of assessment data, and a network of peer schools engaged in professional development. A large cluster randomized trial (West, Morton, & Herlihy, 2016) in third to fifth grades found significantly negative effects after 2 years of intervention (*ES* = -0.09, $p < .05$).

***Acuity*** is an interim assessment program for grade 3 to 8. The program provides online

assessments given to students three times a year to predict their performance on state tests. Teachers are able to intervene early at the student, class, or school level to improve any deficiencies. Across the two cluster randomized evaluations of this method, the weighted mean effect size was +0.16. Konstantopoulos, Miller, and van der Ploeg (2013) found significant positive effects (*ES* = +0.19, $p < .05$), but a later study by Konstantopoulos et al. (2016) found non-significant effects (*ES* = +0.13, *ns*).

***mClass*** is an interim assessment program for grades K to 2. Assessments are conducted face-to-face, where students and teachers work together. The results are entered onto a computer database by the teacher. A cluster randomized trial by Konstantopoulos et al. (2016) with K to 2nd graders found significant negative effects (*ES* = -0.22, $p < .05$)

**Moderator Analyses**

Random-effects models were used to carry out moderator analyses, which identify factors that contribute to positive outcomes (Table 10). We included in these analyses all the qualifying studies except for tutoring studies, as tutoring is so different from other interventions, and only affects a small number of students in each school.

**Research design.** As reported in previous studies, effect sizes may vary according to research design. Cheung and Slavin (2016) found that quasi-experiments produce a significantly higher effect size than randomized studies, on average. We compared effect sizes between randomized trials (*k* = 65, *ES* = +0.08) and quasi-experiments (*k* = 13, *ES* = +0.16), and found a two-to-one ratio favoring quasi-experiments ($Q_m = 2.49$, *ns*).

**Grade levels.** An article by Hill, Bloom, Black, & Lipsey (2007) is often cited to support an expectation that effect sizes will be larger in the early grades, because normal fall-to-spring gains are higher in the early grades. Yet in the present review, this difference

was not seen. To determine if different grade levels may be a source of variation, we divided the studies into those that took place in K to 2 or in 3 to 6. When a study involved students in both categories we divided the outcomes by grades. For five studies the division was not possible. The mean effect size for K-2 studies ($k$ = 22, $ES$ = +0.07) was very similar to the mean effect size for 3-6 studies ($k$ = 40, $ES$ = +0.05) ($Q_m$ = 0.19, *ns*).

**Low-achieving students.** Nine studies reported a separate analysis for low-achieving students, compared to students of middle and high abilities. Note that this analysis only included studies that reported separate analysis for low achievers, so schools that served mostly low achievers (but did not show separate effects for these students) were not included. Mean effect sizes were higher for the low achievers, but this difference was not statistically significant (low achievers: $k$ = 8, $ES$ = +0.08; moderate and high achievers: $k$ = 8, $ES$ = +0.03) ($Q_m$ = 0.89, *ns*).

**Socio-economic status (SES).** To categorize low SES we identified schools with at least 60% of students receiving free lunch. Nine studies did not report SES information and were excluded in this analysis. The effect sizes for low and moderate/high SES were +0.08 ($k$ = 24) and +0.05 ($k$ = 29), respectively. The difference was not statistically significant ($Q_m$ = 0.63, *ns*).

**English language learners (ELLs).** Few studies reported information about ELL. Mean effect sizes for studies involving at least 60% ELLs ($k$ = 7, $ES$ = +0.09) was similar to the mean effect size for studies with lower percentages of ELLs ($k$ = 24, $ES$ = +0.05) ($Q_m$ = 0.73, *ns*).

## Programs Meeting ESSA Standards for Strong and Moderate Evidence of Effectiveness

In 2015, the Every Student Succeeds Act (ESSA) encouraged the use of practices

supported by rigorous evaluations and distinguished different levels of evidence. As previously noted, the "Strong" level requires at least one well-designed and well-implemented randomized study with positive and significant outcomes, and no significant negative effects, and "Moderate" requires at least one quasi-experimental study with positive and significant outcomes and no negative effects.

Table 11 lists the programs that met these ESSA categories, along with the numbers of studies, weighted mean effect sizes, and ESSA ratings. According to ESSA evidence standards, 17 programs met the strong level and 2 programs met the moderate level. Note that these proven programs appear in many categories, not just in those with positive outcomes across all programs. This suggests that there are factors included in particular programs or particular studies that contribute to program effectiveness even if the programs appear to resemble others with less positive outcomes.

## Discussion

This review of evaluations of elementary mathematics programs found 78 studies of very high methodological quality. The studies were mostly randomized and large-scale, increasing the likelihood that their findings will replicate in large-scale applications in practice. Collectively, the studies found that it matters a great deal which programs and which types of programs elementary schools use to teach mathematics. Not surprisingly, one-to-one tutoring by face-to-face adult tutors (*ES* = +0.26) and one-to-small group tutoring (*ES* = +0.32) were particularly effective. It was interesting to find that one-to-one and one-to-small group tutoring did not differ in effectiveness from each other, and that teachers and paraprofessionals were equally effective as tutors. In contrast, on-line tutors and cross-age peer tutors did not show promising impacts. Tutoring has been very successful in elementary reading (Inns et al., 2018a;

Slavin et al., 2011; Wanzek et al., 2016), so it is logical that it would also be effective in math. The findings suggesting no differences between tutoring by paraprofessionals and tutoring by teachers, and between one-to-one and small group tutoring, may indicate that tutoring (by paraprofessionals to small groups) could be a very cost-effective service for students struggling in math. Research on tutoring in reading (Inns et al., 2018a) also found no differences in outcomes between certified teachers and paraprofessionals, but did find that one-to-one tutoring was more effective than one-to-small group.

Programs incorporating technology had variable, but mostly small positive impacts. The average effect size was +0.07. The findings of the review did not provide clear support for any particular approach to technology applications. One study (Gatti, 2013) of SuccessMaker, a widespread computer-assisted instruction approach, found positive outcomes (ES=+0.33), but two others reported no significant differences. There were significant positive effects of a U.K. program, Mathematics and Reasoning (ES=+0.20; Worth et al., 2015). A study of Dream Box (ES=+0.11; Wang & Woodowrth, 2017) found positive effects on geometry achievement, but not on overall math or four other scales. Promising but non-significant positive effects were reported for Time to Know. These programs may well show significant differences in larger studies in the future.

Other than tutoring, the category with the largest effect size (*ES*=+0.25) was instructional process programs, professional development designed to help teachers implement innovative forms of classroom organization and management. Among five studies of four diverse programs, only one, PAX Good Behavior Game, showed significantly positive math outcomes at the cluster level, but one study of TAI and a study of ReflectEd had non-significant but notably positive impacts, and three additional studies of TAI from the 1980s

reported significant positive effects at the cluster level. In the Slavin & Lake (2008) review, instructional process programs had the largest average effect size, and it is therefore interesting to see that these programs still had relatively positive impacts, exceeded only by tutoring (which had not been studied in time to appear in the 2008 review). The effects for all five studies taken together were significant ($p < .01$).

The discrepancy in outcomes was striking between studies of professional development focused on building teachers' knowledge of math content and pedagogy and those of professional development focused on helping teachers implement innovations in classroom organization and management.

Many of the studies of professional development strategies focused on math content and pedagogy used methods that seemed very likely to be effective but were not. One extraordinary example is a study of Intel Math, which provided 80 hours of inservice during the summer to teachers of grades K-8 to improve their understanding of math content and pedagogy. An additional 13 hours were provided to coach teachers based on video tapes of their lessons and to participate in mathematics learning communities to analyze student work. A one-year cluster randomized evaluation with 165 teachers found significantly *negative* impacts on state tests (ES= -0.06, $p < .05$) and nearly identical but non-significant negative effects on NWEA Mathematics. A study of Math Solutions, another professional development approach designed to improve teacher knowledge of math content and pedagogy, found non-significant achievement effect sizes of +0.05 for fourth graders and +0.06 for fifth graders after two years of treatment. A study of a program called Primarily Math (Kutaka et al., 2017) found non-significant effects of a professional development focused on math content and pedagogy. In all three of these studies, there were significant impacts on teachers' knowledge

of mathematics, but this did not transfer to improvement in student achievement. Not one of 12 studies of professional development methods focused on math content and pedagogy achieved statistical significance, and the mean was only +0.04.

The disappointing effects of professional development approaches focused on math content and pedagogy may be due to the fact that programs focused on improving teacher knowledge did not change the daily experience for students very much. It is of course important for teachers to know and apply appropriate math content and pedagogy, but perhaps this is not enough if the student experience is not fundamentally changed.

Beyond tutoring and instructional process models, program outcomes were generally much less positive. A whole-school social-emotional learning program, Positive Action, showed positive math effects in one of two studies (*ES* = +0.16; Snyder et al., 2010). However, other social-emotional approaches did not show positive math achievement outcomes. The overall mean effect size for SEL approaches was +0.03.

Perhaps the most widely used mathematics approaches showed among the smallest impacts. These were mathematics curricula (*ES* = +0.06) and benchmark assessments (*ES* = 0.00). Most of the mathematics curriculum studies just compared a new textbook to existing textbooks, so it is not surprising to see few differences in outcomes. Positive effects were reported in two studies, one of Math Expressions (ES = +0.11; Agodini et al., 2010) and one of Math in Focus (ES = +0.25; ERIA, 2010). For benchmark assessments, the weak outcomes may suggest that teachers' behaviors did not change very much in light of the timely information provided by the interim assessments. However, one of two studies of a benchmark program called Acuity (ES = +0.19; Konstantopoulos, 2013) did find positive outcomes.

Whole-school reform models providing professional development in many subjects,

not just math, had minimal math impacts, on average (*ES* = -0.01). However, the Center for Data-Driven Reform in Education (CDDRE), which helps schools select proven reading and math programs and then implement them, found positive effects on elementary math achievement (*ES* = +0.15; Slavin et al., 2013).

Taken together, the approaches to mathematics education that appeared to have the strongest impacts were ones that strongly emphasize personalization to meet students' needs, and those that emphasize enhancing student engagement and motivation. Tutoring, technology, and instructional process programs, such as cooperative learning and PAX Good Behavior Game, all focus on these attributes. In particular, approaches that give students personal, positive attention from valued adults or other students had the greatest impacts.

As noted earlier, perhaps the most important problems in U.S. math education are the gaps between advantaged and disadvantaged students, and between students of different ethnicities. The evidence from our review has particular bearing on these problems. The approaches with the strongest impacts on math achievement were one-to-one and one-to-small group tutoring. As noted above, the positive outcomes for small group tutoring by paraprofessionals suggest that math tutoring may be an economically feasible way to increase the achievement of low achievers, thereby substantially reducing gaps.

Among approaches other than tutoring, effects were larger for low achievers than for other students, further suggesting pragmatic methods of increasing means while narrowing gaps. Among technology studies in specific, effects were much higher (though not significantly higher, due to small numbers of studies) for both low achievers, compared to students in general, and for schools with at least 60% of student receiving free lunches, compared to other schools. If further research confirms differential effects for technology programs favoring low-

achieving and low-SES students, then it seems likely that using some combination of proven technology approaches and tutoring might significantly reduce gaps in math performance, thereby offering a strategy for ameliorating one of our greatest social and economic as well as educational problems. Further research on means of increasing the effectiveness of both tutoring and technology, and combined strategies using both, would certainly be justified.

One interesting observation about the mathematics programs that did show positive outcomes is that most involved improvements in general pedagogy, not specifically math pedagogy, and in particular on strategies intended to build motivation and positive relationships between teachers and students. For example, the many successful studies of one-to-one and one-to-small group tutoring used tutoring methods adapted from tutoring strategies for reading (Inns et al., 2018). In both subjects, part of the reason for the effectiveness of tutoring is the opportunity it provides for struggling students to form positive relationships with caring adults. PAX Good Behavior Game is designed to increase student motivation and behavior all day, not specifically for math, and Positive Action focuses on social emotional skills in all subjects. These programs had positive impacts in reading as well as math. The Center for Data-Driven Reform in Education also focused equally on reading and math, and found positive outcomes for both subjects. This is not to suggest that improving reading skills is a way to improve math skills, although this may be true to some extent, but rather that the classroom organization and tutoring strategies found to make the most difference in math are effective across the board, perhaps in any subject.

In contrast, few positive outcomes were found for math textbooks, and no significant effects were found for professional development focused on math content and pedagogy. This unexpected pattern of outcomes suggests that successful innovations in math may achieve their

outcomes by improving students' general motivation, social-emotional skills, and behavior, rather than by improving math content, or teachers' math content knowledge or math-specific pedagogy. It is not that content, content knowledge, or pedagogy are unimportant, but that programs focused on these elements may be too similar in focus to what teachers are already doing. In particular, studies of math textbooks invariably compare one (new) text to existing texts, which may not provide enough of a contrast.

If this pattern of findings holds up in future research, it at least suggests that motivation, social-emotional skills, and behavior should be a part of math improvement, along with math content and pedagogy. This may be especially important for students who have a history of failure in math, who may have a particular need to be motivated to learn and enjoy mathematics.

## References

*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.

American Institutes for Research. (2006). *A review of the literature in adult numeracy: research and conceptual issues*. Washington, DC: American Institutes for Research. Retrieved from https://files.eric.ed.gov/fulltext/ED495456.pdf

Ball, D. L., & Cohen, D. K. (1999). Developing practices, developing practitioners: Toward a practice-based theory of professional development. In G. Sykes & L. Darling-Hammond (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 30–32). San Francisco, CA: Jossey-Bass.

Ball, L. D., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*(5), 389–407. doi:10.1177/0022487108324554

Barton, P. E., & Coley, R. J. (2010). *The black-white achievement gap: When progress stopped. Policy information report*. Princeton, NJ: Educational Testing Service.

*Bavarian, N., Lewis, K. M., DuBois, D. L., Acock, A., Vuchinich, S., Silverthorn, N.,…& Flay, B. R. (2013). Using social emotional and character development to improve academic outcomes: A matched pair, cluster randomized controlled trial in low income, urban schools. *Journal of School Health, 83*(11), 771–779. doi:10.1111/josh.12093

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to*

*meta-analysis*. Chichester, UK: John Wiley & Sons.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73*(2), 125–230. doi:10.3102/00346543073002125

Carbonneau, K. J., Marley, S. C., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology, 105*(2), 380–400. doi:10.1111/j.1750-8606

*Cavalluzzo, L., Geraghty, T. M., Steele, J. L., Holian, L., Jenkins, F., Alexander, J. M., & Yamasaki, K. Y. (2014). *"Using data" to inform decisions: How teachers use data to inform practice and improve student performance in mathematics. Results from a randomized experiment of program efficacy*. Arlington, VA: CNA Corporation.

Cheung, A., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review, 9,* 88–113. doi:10.1016/j.edurev.2013.01.001

Cheung, A., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher, 45*(5), 283–292. doi:10.3102/0013189X16656615

Chow, J. C., & Ekholm, E. (2018). Do published studies yield larger effect sizes than unpublished studies in education and special education? A meta-review. *Educational Psychology Review, 30*(3), 727–744. doi:10.1007/s1064801894377

*Clarke, B., Baker, S., Smolkowski, K., Doabler, C., Strand Cary, M., & Fien, H. (2015). Investigating the efficacy of a core kindergarten mathematics curriculum to

improve student mathematics learning outcomes. *Journal of Research on Educational Effectiveness, 8*(3), 303–324. doi:10.1080/19345747.2014.980021

*Clarke, B., Doabler, C.T., Smolkowski, K., Baker, S.K., Fien, H., & Strand Cary, M. (2016). Examining the efficacy of a Tier 2 kindergarten intervention. *Journal of Learning Disabilities, 49*, 152–165. doi:10.1177/0022219414538514

*Clarke, B., Doabler, C. T., Strand Cary, M., Kosty, D., Baker, S., Fien, H., & Smolkowski, K. (2014). Preliminary evaluation of a tier 2 mathematics intervention for first-grade students: Using a theory of change to guide formative evaluation activities. *School Psychology Review, 43*(2), 160–178. Retrieved from https://files.eric.ed.gov/fulltext/ED567775.pdf

Cohen, D. K., & Hill, H. C. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record, 102*(2), 294–343. Retrieved from http://www-personal.umich.edu/~dkcohen/cohen_hill_2000_TCR.pdf

Connor, C. M., Mazzocco, M. M., Kurz, T., Crowe, E. C., Tighe, E. L., Wood, T. S., & Morrison, F. J. (2018). Using assessment to individualize early mathematics instruction. *Journal of School Psychology*, *66*, 97–113. doi: 10.1016/j.jsp.2017.04.005

Corcoran, R., Cheung, A., Kim, E., & Xie, C. (2017). Effective universal school-based social and emotional learning programs for improving academic achievement: A systematic review and meta-analysis of 50 years of research. *Educational Research Review*. doi:10.1016/j.edurev.2017.12.001

de Boer, H., Donker, A. S., & van der Werf, M. P. (2014). Effects of the attributes of educational interventions on students' academic performance: A meta-analysis. *Review*

*of Educational Research*, *84*(4), 509–545. doi:10.3102/0034654314540006

Desimone, L. M., & Garet, M. S. (2015). Best practices in teachers' professional development in the United States. *Psychology, Society and Education*, *7*(3), 252–263. Retrieved from http://repositorio.ual.es/bitstream/handle/10835/3930/Desimone%20En%20ingles.pdf?sequence=1

*Diperna, J. C., Lei, P., Bellinger, J., & Cheng, W. (2016). Effects of a universal positive classroom behavior program on student learning. *Psychology in the Schools, 53*(2), 189–203. doi:10.1002/pits.21891.

*DiPerna, J. C., Lei, P., Cheng, W., Hart, S. C., & Bellinger, J. (2018). A cluster randomized trial of the Social Skills Improvement System-Classwide Intervention Program (SSIS- CIP) in first grade. *Journal of Educational Psychology, 110*(1), 1–16. doi:10.1037/edu0000191

*Doabler, C. T., Clarke, B., Kosty, D. B., Kurtz-Nelson, E., Fien, H., Smolkowski, K., & Baker, S. K. (2016). Testing the efficacy of a Tier 2 mathematics intervention. A conceptual replication study. *Exceptional Children*, *83*(1), 92–110. doi:10.1177/0014402916660084

*Dominguez, P. S., Nicholls, C., & Storandt, B. (2006). *Experimental methods and results in a study of PBS TeacherLine Math Courses*. Syracuse, NY: Hezel Associates. Retrieved from https://files.eric.ed.gov/fulltext/ED510045.pdf

Domitrovich, C. E., Durlak, J. A., Staley, K.C., & Weissberg, R. P. (2017). Social-emotional competence: An essential factor for promoting positive adjustment and reducing risk in school children. *Child Development, 88*, 408–416. doi:10.1111/cdev.12739

*Eddy, R. M., Hankel, N., Hunt, A., Goldman, A., & Murphy, K. (2014). *Houghton Mifflin Harcourt GO Math! efficacy study year two final report*. La Verne, CA: Cobblestone Applied Research & Evaluation, Inc. Retrieved from https://hmhco-v1.prod.webpr.hmhco.com/~/media/sites/home/educators/education-topics/hmh-efficacy/hmh_go_math_rct_yr1_2014.pdf?la=en

*Educational Research Institute of America (2010). *A study of the Singapore math program, Math in Focus, state test results* (Report Number 404). Houghton Mifflin Harcourt.

*Educational Research Institute of America (2013). *A study of the instructional effectiveness of Math in Focus* (Report Number 466). Houghton Mifflin Harcourt.

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). *Teaching adolescents to become learners. The role of noncognitive factors in shaping school performance: A critical literature review.* Chicago, IL: University of Chicago Consortium on Chicago School Research. Retrieved from https://files.eric.ed.gov/fulltext/ED542543.pdf

*Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Schatschneider, C., Hamlett, C. L.,...Bryant, J. D. (2013). Effects of first-grade number knowledge tutoring with contrasting forms of practice. *Journal of Educational Psychology, 105*(1), 58–77. doi:10.1037/a0030127

*Fuchs, L. S., Malone, A. S., Schumacher, R. F., Namkung, J., Hamlett, C. L., Jordan, N. C.,...Changas, P. (2016a). Supported self-explaining during fraction intervention. *Journal of Educational Psychology, 108*(4), 493–508. doi:10.1037/edu0000073

*Fuchs, L. S., Powell, S. R., Seethaler, P. M., Cirino, P. T., Fletcher, J. M., Fuchs, D., &

Hamlett, C. L. (2010). The effects of strategic counting instruction, with and without deliberate practice, on number combination skill among students with mathematics difficulties. *Learning and Individual Differences, 20*(2), 89–100. doi:10.1016/j.lindif.2009.09.003

*Fuchs, L. S., Sterba, S. K., Fuchs, D., & Malone, A. S. (2016b). Does evidence-based fractions intervention address the needs of very low-performing students?. *Journal of Research on Educational Effectiveness, 9*(4), 662–677. doi:10.1080/19345747.2015.1123336

Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: A meta-analysis of randomized controlled trials. *PloSone, 9*(2), e89900. doi:10.1371/journal.pone.0089900

Ganley, C. M., & Lubienski, S. T. (2016). Mathematics confidence, interest, and performance: Examining gender patterns and reciprocal relations. *Learning and Individual Differences, 47*, 182–193. doi:10.1016/j.lindif.2016.01.002

*Garet, M. S., Heppen, J. B., Walters, K., Parkinson, J., Smith, T. M., Song, M., & Borman, G. D. (2016). *Focusing on mathematical knowledge: The impact of content-intensive teacher professional development* (NCEE 2016-4010). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

*Gatti, G. G. (2009). *Pearson SuccessMaker math pilot study. 2008-09 final report*. Pittsburgh, PA: Gatti Evaluation Inc.

*Gatti, G. (2013). *Pearson SuccessMaker response to intervention study: Final report*. Pittsburgh, PA: Gatti Evaluation. Inc. Retrieved from https://www.pearsoned.com/wp-

content/uploads/SM-RTI-Study-old.pdf

*Gatti, G., & Giordano, K. (2008). *Pearson Investigations in Number, Data, & Space efficacy study: 2007-08 School Year Report*. Pittsburgh, PA: Gatti Evaluation, Inc.

*Gatti, G. G., & Petrochenkov, K. (2010). *Pearson SuccessMaker math efficacy study: 2009–10 final report*. Pittsburgh, PA: Gatti Evaluation Inc. Retrieved from https://www.pearsoned.com/wp-content/uploads/successmaker-math-efficacy-report-final.pdf

*Gersten, R., Rolfhus, E., Clarke, B., Decker, L. E., Wilkins, C., & Dimino, J. (2015). Intervention for first graders with limited number knowledge large-scale replication of a randomized controlled trial. *American Educational Research Journal, 52*(3), 516–546. doi:10.3102/0002831214565787

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10), 3–8.

*Glazerman, S., & Seifullah, A. (2012). *An evaluation of the Chicago Teacher Advancement Program (Chicago TAP) after four years. Final report*. Washington, DC: Mathematica Policy Research, Inc. Retrieved from https://www.mathematica-mpr.com/our-publications-and-findings/publications/an-evaluation-of-the-chicago-teacher-advancement-program-chicago-tap-after-four-years

*Gorard, S., Siddiqui, N., & See, B. H. (2015). *Philosophy for Children. Evaluation report and executive summary*. London: Education Endowment Foundation. Retrieved from https://v1.educationendowmentfoundation.org.uk/uploads/pdf/Philosophy_for_Children.pdf

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, *32*(4), 341–370. doi:10.3102/1076998606298043

*Heller, J. I. (2010). *The impact of Math Pathways & Pitfalls on students' mathematics achievement and mathematical language development: A study conducted in schools with high concentrations of Latino/a students and English learners*. San Francisco, CA: WestEd. Retrieved from https://mpp.wested.org/wp-content/uploads/2013/05/mpp-ies-report.pdf

Henderson, S., Petrosino, A., Guckenburg, S., & Hamilton, S. (2007). *Measuring how benchmark assessments affect student achievement* (Issues & Answers Report, REL 2007–No. 039). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast and Islands.

Herman, J. L., Osmundson, E., & Dietel, R. (2010). *Benchmark assessments for improved learning* (AACC Policy Brief). Los Angeles, CA: University of California. Retrieved from https://files.eric.ed.gov/fulltext/ED524108.pdf

Hill, C., Bloom, H., Black, A.R., & Lipsey, M.W. (2007). *Empirical benchmarks for interpreting effect sizes in research.* New York: MDRC.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*(2), 371–406. doi:10.3102/00028312042002371

Inns, A., Lake, C., Pellegrini, M., & Slavin, R. E. (2018a). *Effective programs for struggling readers: A best-evidence synthesis*. Paper presented at the Annual Meeting of the Society for Research on Effective Education, Washington, DC.

Inns, A., Pellegrini, M., Lake, C., & Slavin, R. E. (2018b). *Do small studies add up in the What Works Clearinghouse?* Paper presented at the Annual Meeting of the American Psychological Association, San Francisco.

*Jaciw, A. P., Hegseth, W. M., Lin, L., Toby, M., Newman, D., Ma, B., & Zacamy, J. (2016). Assessing impacts of Math in Focus, a "Singapore Math" program. *Journal of Research on Educational Effectiveness, 9*(4), 473–502. doi:10.1080/19345747.2016.1164777

*Jacob, R., Goddard, R., Kim, M., Miller, R., & Goddard, Y. (2015). Exploring the causal impact of the McREL Balanced Leadership Program on leadership, principal efficacy, instructional climate, educator turnover, and student achievement. *Educational Evaluation and Policy Analysis, 37*(3), 314–332. doi:10.3102/0162373714549620

*Jacob, R., Hill, H., & Corey, D. (2017). The impact of a professional development program on teachers' mathematical knowledge for teaching, instruction, and student achievement. *Journal of Research on Educational Effectiveness, 10* (2), 379–407. doi: 10.1080/19345747.2016.1273411

Jacobse, A. E., & Harskamp, E. G. (2011). *A meta-analysis of the effects of instructional interventions on students' mathematics achievement*. Groningen: GION, Gronings Instituut voor Onderzoek van Onderwijs, Opvoeding en Ontwikkeling, Rijksuniversiteit Groningen. Retrieved from https://www.rug.nl/research/portal/files/10463070/A_Meta-Analysis_of_the_Effects_1.pdf

*Jordan, J. (2009). *Math Connects: National field study: Student learning, student attitudes*

*and teachers' reports on program effectiveness: Evaluation report*. Cincinnati, OH: University of Cincinnati Evaluation Services Center.

*Karper, J., & Melnick, S. A. (1993). The effectiveness of Team Accelerated Instruction on high achievers in mathematics. *Journal of Instructional Psychology*, *20*(1), 49–54.

*Konstantopoulos, S., Miller, S. R., & van der Ploeg, A. (2013). The impact of Indiana's system of interim assessments on mathematics and reading achievement. *Educational Evaluation and Policy Analysis, 35*(4), 481–499. doi:10.3102/0162373713498930

*Konstantopoulos, S., Miller, S. R., van der Ploeg, A., & Li, W. (2016). Effects of interim assessments on student achievement: Evidence from a large-scale experiment. *Journal of Research on Educational Effectiveness, 9*(sup1), 188–208. doi:10.1080/19345747.2015.1116031

*Kutaka, T. S., Smith, W. M., Albano, A. D., Edwards, C. P., Ren, L., Beattie, H. L.,...Stroup, W. W. (2017). Connecting teacher professional development and student mathematics achievement: Mediating belonging with multimodal explorations in language, identity, and culture. *Journal of Teacher Education, 68*(2), 140–154. doi:10.1177/0022487116687551

*Lambert, R., Algozzine, B., & McGee, J. (2014). Effects of progress monitoring on math performance of at-risk students. *British Journal of Education, Society and Behavioural Science, 4*(4), 527–540. Retrieved from http://www.journalrepository.org/media/journals/BJESBS_21/2014/Jan/Lambert442013BJESBS7259_1.pdf

*Lehmann, R. H., & Seeber, S. (2005). *Accelerated Math in grades 4 through 6: Evaluation of an experimental program in 15 schools in North Rhine-Westphalia*. Berlin:

Humboldt University.

Li, Q., & Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. *Educational Psychology Review, 22*(3), 215–243. doi:10.1007/s10648-010-9125-8

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage

*Lloyd, C., Edovald, T., Morris, S., Kiss, Z., Skipp, A., & Haywood, S. (2015). *Durham shared maths project. Evaluation report and Executive summary.* London: Education Endowment Foundation. Retrieved from https://v1.educationendowmentfoundation.org.uk/uploads/pdf/Shared_Maths_1.pdf

*Magnolia Consulting (2012). *A final report for the evaluation of Pearson's Waterford Early Learning program: Year 2*. Charlottesville, VA: Magnolia Consulting.

*Manchester Institute of Education (2015). *Promoting Alternative Thinking Strategies (PATHS). Evaluation report and executive summary*. London: Education Endowment Foundation. Retrieved from https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/EEF_Project_Report_PromotingAlternativeThinkingStrategies.pdf

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, NY: Cambridge University Press.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med, 6*(7), e1000097. doi:10.1371/journal.pmed1000097

*Motteram, G., Choudry, S., Kalambouka, A., Hutcheson, G., & Barton, A. (2016). *ReflectED. Evaluation report and executive summary*. London: Education Endowment Foundation.

National Center for Educational Statistics (2018). *National Assessment of Educational*

*Progress.* Washington, DC: National Center for Educational Statistics.

National Council of Teachers of Mathematics (2014). *Principles to actions: Ensuring mathematical success for all*. Reston, VA: National Council of Teachers of Mathematics.

National Research Council. (2004). *On evaluating curricular effectiveness: Judging the quality of K-12 mathematics evaluations*. Washington, DC: The National Academies Press. doi:10.17226/11025

*Newman, D., Finney, P. B., Bell, S., Turner, H., Jaciw, A. P., Zacamy, J. L., & Gould, L. F. (2012). *Evaluation of the effectiveness of the Alabama Math, Science, and Technology Initiative (AMSTI).* (NCEE 2012–4008). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

*O'Connor, E. E., Cappella, E., McCormick, M. P., & McClowry, S. G. (2014). An examination of the efficacy of INSIGHTS in enhancing the academic and behavioral development of children in early grades. *Journal of Educational Psychology, 106*(4), 1156–1169. doi:10.1037/a0036615

Pellegrini, M., Inns, A., Lake, C., & Slavin, R. E. (2018). *Effects of types of measures on What Works Clearinghouse outcomes.* Paper presented at the Annual Meeting of the American Psychological Association, San Francisco.

R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/.

*Randel, B., Apthorp, H., Beesley, D., Clark, F., & Wang, X. (2016). Impacts of professional development in classroom assessment on teacher and student outcomes.

*The Journal of Educational Research, 109*(5), 491–502. doi:10.1080/00220671.2014.992581

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

*Reid, E. E., Chen, J. Q., & McCray, J. (2014). *Achieving high standards for Pre-K-Grade 3 mathematics: A whole teacher approach to professional development*. Paper presented at the Annual Meeting of the Society for Research on Effective Education, Washington, DC.

*Resendez, M., & Azin, M. (2006). *2005 Scott Foresman–Addison Wesley Elementary Math randomized control trial: Final report*. Jackson, WY: PRES Associates, Inc.

*Resendez, M., Azin, M., & Strobel, A. (2009). *A study on the effects of Pearson's 2009 enVision math program: Final summative report*. Jackson, WY: Press Associates. Retrieved from https://www.pearsoned.com/wp-content/uploads/envisionmath-efficacy-report-year-2.pdf

*Resendez, M., & Manley, M. A. (2005). *Final report: A study on the effectiveness of the 2004 Scott Foresman–Addison Wesley Elementary Math program*. Jackson, WY: PRES Associates.

*Rimm-Kaufman, S. E., Fan, X., Chiu, Y. J., & You, W. (2007). The contribution of the Responsive Classroom approach on children's academic achievement: Results from a three year longitudinal study. *Journal of School Psychology, 45*(4), 401–421. doi:10.1016/j.jsp.2006.10.003

*Rimm-Kaufman, S. E., Larsen, R. A. A., Baroody, A. E., Curby, T. W., Ko, M., Thomas, J. B.,…DeCoster, J. (2014). Efficacy of the "Responsive Classroom" approach:

Results from a 3-year, longitudinal randomized controlled trial. *American Educational Research Journal, 51*(3), 567–603. doi:10.3102/0002831214523821

*Rosen, Y., & Beck-Hill, D. (2012). Intertwining digital content and a one-to-one laptop environment in teaching and learning: Lessons from the Time to Know program. *Journal of Research on Technology in Education, 44*(3), 225–241. doi:10.1080/15391523.2012.10782588

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2006). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. West Sussex, UK: John Wiley & Sons.

*Rutherford, T., Farkas, G., Duncan, G., Burchinal, M., Kibrick, M., Graham, J.,…Martinez, E. (2014). A randomized trial of an elementary school mathematics software intervention: Spatial-Temporal Math. *Journal of Research on Educational Effectiveness, 7*(4), 358–383. doi:10.1080/19345747.2013.856978

*Rutt, S., Easton, C., & Stacey, O. (2014). *Catch Up® Numeracy: Evaluation report and executive summary*. London, UK: Education Endowment Foundation. Retrieved from https://www.nfer.ac.uk/publications/EFCU01/EFCU01.pdf

Savelsbergh, E. R., Prins, G. T., Rietbergen, C., Fechner, S., Vaessen, B. E., Draijer, J. M., & Bakker, A. (2016). Effects of innovative science and mathematics teaching on student attitudes and achievement: A meta-analytic study. *Educational Research Review, 19*, 158–172. doi:10.1016/j.edurev.2016.07.003

Saxe, G. B., & Gearhart, M. (2001). Enhancing students' understanding of mathematics: A study of three contrasting approaches to professional support. *Journal of Mathematics Teacher Education, 4*(1), 55–79. doi:10.1023/A:1009935100676

*Schoen, R. C., LaVenia, M., & Tazar, A. M. (2018, March). *Effects of a two-year Cognitively Guided Instruction professional development program on first- and second-grade student achievement in mathematics.* Paper presented at the annual meeting of the Society for Research in Effective Education, Washington, DC.

*ServeMinnesota (2017). *Research brief: Minnesota Math Corps*. Minneapolis: ServeMinnesota. Retrieved from https://minnesotamathcorps.org/sites/default/files/Math%20Corps%20White%20Paper.pdf

Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher, 15*(9), 5–11. doi:10.3102/0013189X015009005

*Slavin, R. E., Cheung, A., Holmes, G., Madden, N. A., & Chamberlain. A. (2013). Effects of a data-driven reform model on state assessment outcomes. *American Educational Research Journal, 50*(2), 371–396. doi:10.3102/0002831212466909

Slavin, R. E., & Karweit, N. L. (1985). Effects of whole-class, ability grouped and individualized instruction on mathematics achievement. *American Educational Research Journal, 22*(3), 351–367. doi: 10.3102/00028312022003351

Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research, 78*(3), 427–515. doi:10.3102/0034654308317473

Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2011). Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review, 6*(1),

1–26. doi:10.1016/j.edurev.2010.07.002

Slavin, R. E., Leavey, M., & Madden, N.A. (1984). Combining cooperative learning and individualized instruction: Effects on student mathematics achievement, attitudes, and behaviors. *Elementary School Journal, 84*, 409–422.

*Smith, T. M., Cobb, P., Farran, D. C., Cordray, D. S., & Munter, C. (2013). Evaluating Math Recovery: Assessing the causal impact of a diagnostic tutoring program on student achievement. *American Educational Research Journal, 50*(2), 397–428. doi:10.3102/0002831212469045

*Snyder, F., Flay, B., Vuchinich, S., Acock, A., Washburn, I., Beets, M., & Li, K. K. (2010). Impact of a social-emotional and character development program on school-level indicators of academic achievement, absenteeism, and disciplinary outcomes: A matched-pair, cluster-randomized, controlled trial. *Journal of Research on Educational Effectiveness, 3*(1), 26–55. doi:10.1080/19345740903353436

*Solomon, T., Martinussen, R., Dupuis, A., Gervan, S., Chaban, P., Tannock, R., & Ferguson, B. (2011). *Investigation of a cognitive science based approach to mathematics instruction.* Paper presented at the Biennial Meeting of the Society for Research in Child Development, Montreal, Canada.

*Stevens, R. J., & Slavin, R. E. (1995). The cooperative elementary school: Effects on students' achievement, attitudes, and social relations. *American Educational Research Journal, 32*(2), 321–351. doi:10.3102/00028312032002321

*Strobel, A., Resendez, M., & DuBose, D. (2017). *enVisionmath2.0 Year 2 RCT Study Final Report.* Thayne, WY: Strobel Consulting, LLC.

*Styers, M. & Baird-Wilkerson, S. (2011). *A final report for the evaluation of*

*Pearson's focusMATH Program*. Charlottesville, VA: Magnolia Consulting.

*Suppes, P., Holland, P. W., Hu, Y., & Vu, M.T. (2013). Effectiveness of an individualized computer-driven online math K-5 course in eight California Title I elementary schools. *Educational Assessment, 18*(3), 162–181. doi:10.1080/10627197.2013.814516

*Torgerson, C., Ainsworth, H., Buckley, H., Hampden-Thompson, G., Hewitt, C., Humphry, D., Jefferson, L., Mitchell, N., & Torgerson, D. (2016). *Affordable Online Maths Tuition. Evaluation report and executive summary*. London: Education Endowment Foundation. Retrieved from http://dro.dur.ac.uk/19387/1/19387.pdf?DDD29+hsmz78+dul4eg

*Torgerson, C. J., Wiggins, A., Torgerson, D., Ainsworth, H., & Hewitt, C. (2013). Every Child Counts: Testing policy effectiveness using a randomised controlled trial, designed, conducted and reported to CONSORT standards. *Research In Mathematics Education, 15*(2), 141–153. doi:10.1080/14794802.2013.797746

*Vaden-Kiernan, M., Borman, G., Caverly, S., Bell, N., de Castilla, V. R., & Sullivan, K. (2015). *Preliminary findings from a multi-year scale-up effectiveness trial of Everyday Mathematics*. Paper presented at the Society for Research on Effective Education, Washington, DC.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48. doi:10.18637/jss.v036.i03

*Vignoles, A., Jerrim, J., & Cowan, R. (2015). *Mathematics Mastery: Primary evaluation report*. London: Education Endowment Foundation. Retrieved from https://v1.educationendowmentfoundation.org.uk/uploads/pdf/Mathematics_Mastery_Pri

mary_(Final)1.pdf

Vygotsky, L.S. (1978). *Mind in Society* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge, MA: Harvard University Press.

*Wang, H., & Woodworth, K. (2011a). *Evaluation of Rocketship Education's use of DreamBox Learning's online mathematics program.* Menlo Park, CA: SRI International. Retrieved

*Wang, H., & Woodworth, K. (2011b). *A randomized controlled trial of two online mathematics curricula*. Paper presented at the Annual Meeting of the Society for Research on Effective Education, Washington, DC.

Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M. A., & Capin, P. (2016). Meta-analyses of the effects of tier 2 type reading interventions in grades K-3. *Educational Psychology Review*, *28*(3), 551–576. doi:10.1007/s10648-015-9321-7

Waters, J. T., Marzano, R. J., & McNulty, B. (2003). *Balanced leadership: What 30 years of research tells us about the effect of leadership on student achievement*. Aurora, CO: Mid-continent Research for Education and Learning. Retrieved from http://www.peecworks.org/peec/peec_research/I01795EFA.0/Marzano

*Weis, R., Osborne, K. J., & Dean, E. L. (2015). Effectiveness of a universal, interdependent group contingency program on children's academic achievement: A countywide evaluation. *Journal of Applied School Psychology, 31*(3), 199–218. doi:10.1080/15377903.2015.1025322

*West, M. R., Morton, B. A., & Herlihy, C. M. (2016). *Achievement Network's Investing in Innovation expansion: Impacts on educator practice and student achievement*. Cambridge, MA: Center for Educational Policy Research, Harvard University. Retrieved from https://cepr.harvard.edu/files/cepr/files/anet-research-report.pdf

What Works Clearinghouse. (2017). *Procedures handbook* (Version 4.0). Washington, DC: What Works Clearinghouse.

*Wijekumar, K., Hitchcock, J., Turner, H., Lei, P. W., & Peck, K. (2009). *A multisite cluster randomized trial of the effects of CompassLearning Odyssey® Math on the math achievement of selected Grade 4 students in the Mid-Atlantic region* (NCEE 2009-4068). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

*Wilkerson, S. B., Shannon, L. C., Styers, M. K., & Grant, B. (2012). *A study of the effectiveness of a school improvement intervention (Success in Sight).* (NCEE 2012-4014). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

*Worth, J., Sizmur, J., Ager, R., & Styles, B. (2015). *Improving numeracy and literacy*. London: Education Endowment Foundation. Retrieved from https://www.nfer.ac.uk/publications/EEOL01/EEOL01.pdf

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007–No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.

*Ysseldyke, J., & Bolt, D. M. (2007). Effect of technology-enhanced continuous progress monitoring on math achievement. *School Psychology Review, 36*(3), 453–467.

Table 1

*Weighted Mean Effect Sizes of Program Categories*

| *Program Categories* | *k* | *ES* | *95% CI* | | *Q* | $I^2$ | $\tau^2$ |
|---|---|---|---|---|---|---|---|
| | | | *LL* | *UL* | | | |
| *Tutoring Programs* | 15 | +0.25*** | +0.18 | +0.32 | 20.03 | 25.08 | 0.00 |
| One-to-one Tutoring | 6 | +0.26*** | +0.17 | +0.34 | 1.30 | 0.00 | 0.00 |
| Small Group Tutoring | 7 | +0.32*** | +0.21 | +0.43 | 7.80 | 16.04 | 0.00 |
| *Programs Incorporating Technology* | 14 | +0.07* | +0.00 | +0.14 | 12.96 | 23.19 | 0.00 |
| *Professional Development for Math Content and Pedagogy* | 12 | +0.04 | -0.02 | +0.11 | 4.00 | 0.00 | 0.00 |
| *Instructional Process Programs* | 5 | +0.25*** | +0.09 | +0.40 | 1.89 | 0.00 | 0.00 |
| *Whole-School Reform* | 4 | -0.01 | -0.12 | +0.11 | 1.59 | 0.00 | 0.00 |
| *Social-Emotional Interventions* | 8 | +0.03 | -0.10 | +0.15 | 1.90 | 0.00 | 0.00 |
| *Mathematics Curricula* | 16 | +0.06 | 0.00 | +0.12 | 14.98 | 7.77 | 0.00 |
| *Benchmark Assessments* | 4 | 0.00 | -0.18 | +0.17 | 7.16 | 58.92 | 0.02 |

*Note*. *k* = total number of studies; *ES* = effect size; CI = confidence interval; LL = lower limit, UL = upper limit.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 2

*Tutoring Programs*

| *Study* | *Design* | *Duration* | *k* | *Grade* | *Sample Characteristics* | *Posttest* | *ES by Subgroup/ Measure* | *Overall ES* |
|---|---|---|---|---|---|---|---|---|
| | | | | *One-to-one Tutoring by Teachers* | | | | |
| Math Recovery | | | | | | | | |
| Smith et al. (2013) | QE | 1 year | 775 students (259E, 516C) | 1 | 48% minority, 15% ELL, 65% FRL. | WJ-III<br>Math Fluency<br>App. Problems<br>Quant Concepts<br>Math Reasoning | <br>+0.15*<br>+0.28*<br>+0.24*<br>+0.30* | +0.24* |
| Numbers Count | | | | | | | | |
| Torgerson et al. (2013) | SR | 12 weeks | 418 students (144E, 274C) | 6-7 years old (U. S. grade 1) | England. 75% FRL. | Progress in Math (PIM 6) | | +0.33* |
| | | | | *One-to-one Tutoring by Paraprofessionals* | | | | |
| Catch Up® Numeracy | | | | | | | | |
| Rutt et al. (2014) | SR | 30 weeks | 216 students (108E, 108C) | Year 2-6 (U.S. grade 1-5) | England. 35% FRL. | Basic Number Screening Test | | +0.21* |
| Galaxy Math | | | | | | | | |
| Fuchs et al. (2013) | SR | 16 weeks | 591 students (385E, 206C) | 1 | Southeast school district. 69% AA, 7% H, 83% FRL. | Word Problems | | +0.25* |
| Pirate Math | | | | | | | | |
| Fuchs et al. (2010) | SR | 16 weeks | 150 students (100E, 50C) | 3 | Nashville and Houston. 35% SPED, 19% ELL, 56% AA, 29% H. | KeyMath | | +0.37* |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *One-to-one Tutoring Provided by Paid Volunteers* | | | | | | | | |
| MathCorps | | | | | | | | |
| ServeMinnesota (2017) | SR | 6 months | 284 students (183E, 101C) | 4-6 | Minnesota. 35%W, 27%AA, 20% A, 61%FRL. | STAR Math | | +0.20* |
| *One-to-One Tutoring Delivered Online* | | | | | | | | |
| Affordable Online Maths Tuition | | | | | | | | |
| Torgerson et al. (2016) | CR | 27 weeks | 64 schools 578 students (289E, 289C) | Year 6 (U.S. grade 5) | England. 92% FRL, 43% minority. | Key Stage 2 | | -0.03 |
| *Cross-Age Tutoring* | | | | | | | | |
| Shared Maths | | | | | | | | |
| Lloyd et al. (2015) | CR | 2 years | 79 schools Year 3 (tutees) 2786 students Year 5 (tutors) 2683 students | Year 3, 5 (U. S. grades 2, 4) | England. 22% FRL, 86% W, 4% AA, 5% A. | ICAS Year 3 Year 5 | +0.01 +0.02 | +0.02 |
| *One-to-Small Group Tutoring by Teachers* | | | | | | | | |
| Number Rockets | | | | | | | | |
| Gersten et al. (2015) | CR | 6 months | 76 schools 994 students (615E, 379C) | 1 | 44% AA, 46% H, 34% FRL. | TEMA-3 | | +0.34* |
| *One-to-Small Group Tutoring by Paraprofessionals* | | | | | | | | |
| FocusMATH | | | | | | | | |
| Styers & Baird-Wilkerson (2011) | SR | 1 year | 341 students (166E, 175C) | 3, 5 | 23% AA, 33% H, 24% ELL, 12% SPED, 71% FRL | KeyMath 3 | | +0.24* |
| Fraction Face-Off! | | | | | | | | |
| Fuchs et al. (2016a) | SR | 12 weeks | 213 students (143E, 70C) | 4 | Students at risk from 14 schools. | NAEP Items | | +0.39* |
| Fuchs et al. (2016b) | SR | 12 weeks | 212 students (142E, 70C) | 4 | 49% AA, 27% H, 18% ELL, 90% FRL. | NAEP Items | | +0.64* |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Fusion Math | | | | | | | | |
| Clarke et al. (2014) | SR | 19 weeks | 78 students (38E, 40C) | 1 | Pacific Northwest. 20% H, 18% ELL, 70% FRL, 12% SPED. | SAT-10 | | +0.11 |
| ROOTS | | | | | | | | |
| Doabler et al. (2016) | SR | 5 months | 292 students (208E, 82C) | K | Boston. 7% AA, 89% W, 50% H, 26% ELL. | TEMA-3<br>NSB<br>SESAT | +0.31*<br>+0.40*<br>+0.24 | +0.32* |
| Clarke et al. (2016) | SR | 4 months | 290 students (203E, 87C) | K | Oregon. 5% AA, 58% W, 33% H, 32% LEP, 11% SPED | TEMA-3<br>NSB<br>SESAT | +0.32*<br>+0.16<br>+0.001 | +0.16 |

*Note for Tables 2-9.*
Design/Treatment: SR=Student Randomized, CR=Cluster Randomized, QE=Quasi Experiment, CQE=Cluster Quasi-Experiment
Measures: BAM: Balanced Assessment in Mathematics, CAT: California Achievement Test, CMT-Math: Connecticut Mastery Test, CST: California Standards Test, CSAP: Colorado Student Assessment Program, ECLS-K: Early Childhood Longitudinal Program, FCAT: Florida Comprehensive Assessment Test, GMADE: Group Mathematics Assessment and Diagnostic Evaluation, HCPS II: Hawaii Content and Performance Standards, ICAS: Interactive Computerised Assessment Systemin, CAS: Interactive Computerized Assessment System, ISAT: Illinois Student Achievement Test, ISTEP+: Indiana State Test of Educational Proficiency, ITBS: Iowa Test of Basic Skills, MAP: Measure of Academic Progress, MAT- Metropolitan Achievement Test, MEAP: Michigan Educational Assessment Program, NAEP: National Assessment of Educational Progress, NJASK: New Jersey State Test; NSB: Brief Number Sense Screener, Nevada CRT: Nevada Criterion Referenced Test, NWEA: Northwest Evaluation Association, SAT 10: Stanford Achievement Test 10, SESAT: Stanford Early School Achievement Test; SOL: Virginia Standards of Learning, STAR Math: Standardized Testing and Reporting, TAKS: Texas Assessment of Knowledge and Skills, TEMA-3: Test of Early Mathematics Ability 3, WJ III: Woodcock-Johnson III.
Demographics: A=Asian, AA=African-American, H=Hispanic, W=White, FRL=Free/Reduced Lunch, ELL=English Language Learner, LD=Learning Disabilities, SPED=Special Education.
* $p < .05$ at the appropriate level of analysis (cluster or individual).

Table 3
*Programs Incorporating Technology*

| *Study* | *Design* | *Duration* | *k* | *Grade* | *Sample Characteristics* | *Posttest* | *ES by Subgroup/ Measure* | *Overall ES* |
|---|---|---|---|---|---|---|---|---|
| | | | | *Typical CAI* | | | | |
| Accelerated Math | | | | | | | | |
| Lehmann & Seeber (2005) | CQE | 4 months | 47 classes<br>1243 students (577E, 666C) | 4-6 | Germany. approx. 18% immigrants | Hamburger Schulleistungstest | | +0.06 |
| Ysseldyke & Bolt (2007) | CR | 1 year | 36 classes<br>723 students (368E, 355C) | 2-5 | AL, FL, SC, TX, MS, MI, NC. 44% AA, 45% H | TerraNova | | 0.00 |
| Lambert et al. (2014) | CR | 1 year | 36 classes<br>504 students (256E, 248C) | 2-5 | Midwestern US.<br>40% minority, 76% FRL, 18% SPED | TerraNova | | +0.03 |
| DreamBox Learning | | | | | | | | |
| Wang & Woodworth (2011a) | SR | 4 months | 557 students<br>(446E, 111C) | K, 1 | San Francisco Bay Area.<br>87% H, 81% ELL, 88% FRL. | NWEA<br>Math overall<br>Problem solving<br>Number sense<br>Computation<br>Geometry<br>Statistics | <br>+0.11<br>+0.06<br>+0.08<br>+0.13<br>+0.16*<br>+0.12 | +0.11 |
| Educational Program for Gifted Youth (EGPY) | | | | | | | | |
| Suppes et al. (2013) | SR | 1 year | 1484 students<br>(742E, 742C) | 2-5 | California.<br>55% AA, 31% H. | CST | | -0.01 |
| Odyssey Math | | | | | | | | |
| Wijekumar et al. (2009) | CR | 1 year | 122 teachers<br>2,456 students<br>(1,223E, 1,233C) | 4 | DE, NJ, PA. 18% FRL, 25% minority, 7% ELL. | TerraNova | | +0.02 |
| SuccessMaker | | | | | | | | |
| Gatti (2009) | CQE | 1 year | 8 schools<br>792 students (455E, 337C) | 3,5 | AZ, FL, MA, NJ.<br>34% H, 34% FRL, 89% ELL, 47% low achievers. | GMADE<br>Grade 3<br>Grade 5 | <br>+0.11<br>+0.03 | +0.07 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Gatti & Petrochenkov (2010) | CR | 1 year | 47 classes<br>913 students (506E, 407C) | 3, 5 | AZ, AR, CA, IN, KS, PA.<br>88% ELL, 66% FRL,42% H,<br>12% AA, 40% low achievers. | GMADE<br>Grade 3<br>Grade 5 | <br>+0.27<br>-0.19 | <br><br>+0.05 |
| Gatti (2013) | SR | 1 year | 490 students (239E, 251C) | 5 | AZ, CA, KS, MI, OR, TX.<br>49% H, 8% AA,<br>11% SPED, 17% LEP,<br>70% FRL. | GMADE<br>AIMSweb Comp.<br>Conc. and App. | +0.09<br>+0.42*<br>+0.49* | +0.33* |
| Waterford Early Learning | | | | | | | | |
| Magnolia Consulting (2012) | CR | 2 years | 57 classes<br>680 students (425E, 255C) | K-1<br>1-2 | 19% AA, 53% H, 17% W,<br>73% FRL, 32% LEP, 5%<br>SPED. | SAT 10 | | +0.04 |
| | | | | | | | | |
| Mathematics and Reasoning | | | | | | | | |
| Worth et al. (2015) | CR | 4 months | 36 schools<br>1365 students (517E, 848C) | Year 2<br>(grade 1) | England. 16% FRL,<br>14% SPED, 14% ELL. | Progress in Math<br>(PIM 7) | | +0.20* |
| Reasoning Mind | | | | | | | | |
| Wang & Woodworth (2011b) | SR | 4 months | 651 students<br>(521E, 130C) | 2-5 | San Francisco Bay Area.<br>87% H, 81% ELL, 88% FRL. | NWEA<br>Math overall<br>Problem solving<br>Number sense<br>Computation<br>Geometry<br>Statistics | <br>-0.02<br>-0.05<br>+0.01<br>-0.08<br>+0.11<br>-0.02 | -0.01 |
| ST Math | | | | | | | | |
| Rutherford et al. (2014) | CR | 1, 2 years | 1 year: 34 schools<br>10455 students<br>2 years: 18 schools<br>2677 students | 3-5 | Southern CA. 90% FRL,<br>85% H, 63% ELL. | CST<br>1 year<br>2 years | <br>+0.09<br>+0.03 | <br>+0.08 |
| Time to Know | | | | | | | | |
| Rosen & Beck-Hill (2012) | CQE | 6 months | 4 schools<br>476 students (283E, 193C) | 4,5 | Districts in Dallas, TX<br>18% AA, 63% H | TAKS | | +0.31 |

Table 4
*Professional Development for Math Content and Pedagogy*

| *Study* | *Design* | *Duration* | *k* | *Grade* | *Sample Characteristics* | *Posttest* | *ES by Subgroup/ Measure* | *Overall ES* |
|---|---|---|---|---|---|---|---|---|
| Intel Math | | | | | | | | |
| Garet et al. (2016) | CR | 1 year | 165 teachers<br>3677 students<br>(1760E, 1917C) | 4 | 46% W, 14% AA, 30% H, 58% FRL, 12% ELL, 14% SPED. | State tests<br>NWEA | -0.06*<br>-0.05 | -0.06* |
| CASL | | | | | | | | |
| Randel et al. (2016) | CR | 1-2 years | 67 schools<br>9,596 students<br>(4,420E, 5,176C) | 4,5 | CO.<br>56% W, 27% H, 47% FRL. | CSAP | | +0.01 |
| Using Data | | | | | | | | |
| Cavalluzzo et al. (2014) | CR | 2 years | 59 schools<br>10,877 students<br>(5,384E, 4,903C) | 4,5 | FL.<br>47% AA, 9% H, 66% FRL, 10% SPED. | FCAT | | +0.01 |
| Math Solutions | | | | | | | | |
| Jacob et al. (2017) | CR | 3 years | 74 classes<br>1453 students<br>(727E, 726C) | 4, 5 | 63% AA, 21% W, 14% Sped | State tests | | +0.06 |
| Cognitively Guided Instruction | | | | | | | | |
| Schoen et al. (2018) | CR | 2 years | 22 schools (11 E, 11C)<br>2230 students (1110 E, 1120C) | 1, 2 | 37%W, 37% H, 18% AA, 22% ELL, 60% FRL | ITBS<br>Grade 1<br>Computations<br>Problems<br>Grade 2<br>Computations<br>Problems | <br><br>-0.08<br>0.09<br><br>-0.07<br>0.06 | 0.00 |
| AMSTI | | | | | | | | |
| Newman et al. (2012) | CR | 1 year | 40 schools<br>9,370 students<br>(5,111E, 4,259C) | 4-5 | 49% minority, 64% FRL. | SAT 10 | | +0.05 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| EarlyMath | | | | | | | | |
| Reid et al. (2014) | CQE | 2 years | 16 schools<br>903 students (443, 460C) | K-2 | Schools in a large Midwestern city. | W-J III Applied Problems | | +0.01 |
| Math Pathways & Pitfalls | | | | | | | | |
| Heller (2010) | CR | 1 year | 121 classes<br>2,160 students (1,204E, 956C) | 4, 5 | AZ, CA, IL.<br>55% ELL, 76% FL, 8% AA, 69% H, 9% W. | State tests<br>Grade 4<br>Grade 5 | <br>+0.04<br>+0.08 | +0.06 |
| Mathematics Mastery | | | | | | | | |
| Vignoles et al. (2015) | CR | 1 year | 83 schools<br>4,176 students (2,160E, 2,016C) | Year 1 (U. S. grade K) | Schools across England. | Number Knowledge Test | | +0.10 |
| PBS TeacherLine | | | | | | | | |
| Dominguez et al. (2006) | CR | 1 year | 87 teachers<br>1,119 students (523E, 596C) | 3-5 | FL, SC, NY. | Algebra test<br>Geometry test | -0.02<br>+0.08 | +0.03 |
| Philosophy for Children | | | | | | | | |
| Gorard (2015) | CR | 1 year | 48 schools<br>1529 students (772E, 757C) | Year 5 (U.S. grade 4) | England. 47% FRL, 19% SPED, 12% ELL, 26% minority. | Key Stage 2 | | +0.10 |
| Primarily Math | | | | | | | | |
| Kutaka et al. (2017) | CQE | 1 year | 218 teachers<br>809 students (313E, 496C) | K-2 | 3 urban school districts. | TEMA-3 | | +0.14 |

Table 5
*Instructional Process Programs*

| *Study* | *Design* | *Duration* | *k* | *Grade* | *Sample Characteristics* | *Posttest* | *ES by Subgroup/ Measure* | *Overall ES* |
|---|---|---|---|---|---|---|---|---|
| TAI | | | | | | | | |
| Stevens & Slavin (1995) | CQE | 2 years | 5 schools<br>873 students (411E, 462C) | 2-6 | MD. 7% minority, 10% FRL, 9% SPED. | CAT<br>Computation<br>Application | +0.29<br>+0.10 | +0.20 |
| Karper & Melnick (1993) | CQE | 1 year | 8 classes<br>165 students (84E, 81C) | 4-5 | Hershey, PA. | District Test<br>Grade 4<br>Grade 5 | <br>-0.05<br>-0.12 | -0.09 |
| PAX Good Behavior Game | | | | | | | | |
| Weis et al. (2015) | CQE | 1 year | 49 classes<br>703 students (402E, 301C) | 1,2 | Ohio. 82% W, 48% FRL. | MAP | | +0.32* |
| Individualized Student Instruction (ISI) | | | | | | | | |
| Connor et al. (2018) | CR | 1 year | 5 schools<br>32 teachers<br>370 students<br>(205E, 165C) | 2 | North FL. 84%W, 5% AA | Woodstock Math<br>Fluency<br>Key Math | <br>+0.16<br>+0.07 | +0.12 |
| ReflectEd | | | | | | | | |
| Motteram et al. (2016) | CR | 1 year | 65 classes<br>1570 students (839E, 731C) | Year 5<br>(U.S. grade 4) | England | InCAS | | +0.30 |

Table 6
*Whole-School Reform Programs*

| *Study* | *Design* | *Duration* | *k* | *Grade* | *Sample Characteristics* | *Posttest* | *ES by Subgroup/ Measure* | *Overall ES* |
|---|---|---|---|---|---|---|---|---|
| Center for Data-Driven Reform in Education (CDDRE) | | | | | | | | |
| Slavin et al. (2013) | CR | 4 years | 20 districts<br>11,484 students<br>(5,742E, 5,742C) | 5 | PA, AZ, MS, IN, OH, TN, AL.<br>64% FRL, 29% AA, 20% H,<br>48% W. | State Tests | | +0.15* |
| McREL Balanced Leadership | | | | | | | | |
| Jacob et al. (2015) | CR | 3 years | 119 schools<br>21,420 students | 3-5 | MI. 47% FRL, 90% W. | MEAP | | +0.03 |
| Success in Sight | | | | | | | | |
| Wilkerson et al. (2012) | CR | 2 years | 52 schools<br>8,213 students<br>(4,413E, 3,800C) | 3-5 | MN, MO. 40% W, 33% AA,<br>10% H, 16% A, 70% FRL. | State tests | | -0.11* |
| The System for Teacher and Student Achievement (TAP) | | | | | | | | |
| Glazerman & Seifullah (2012) | CR | 1 year | 34 schools<br>4,588 students<br>(2,294E, 2,294C) | 4-6 | Chicago, IL.<br>91% AA, 96% FRL, 14%<br>SPED, 8% H, 3% ELL. | ISAT<br>Grade 4<br>Grade 5<br>Grade 6 | <br>-0.07<br>-0.11<br>+0.01 | -0.06 |

Table 7
*Social-Emotional Learning Programs*

| *Study* | *Design* | *Duration* | *k* | *Grade* | *Sample Characteristics* | *Posttest* | *ES by Subgroup/ Measure* | *Overall ES* |
|---|---|---|---|---|---|---|---|---|
| INSIGHTS | | | | | | | | |
| O'Connor et al. (2014) | CR | 1, 2 years | 22 schools<br>435 students (225E, 210C) | K-1 | 75% AA, 16%H, 87% FRL. | WJ-III App. Prob. | | +0.04 |
| Positive Action | | | | | | | | |
| Snyder et al. (2010) | CR | 4 years | 20 schools<br>10880 students (5,440E, 5,440C) | 5,6 | 5% AA, 14% Filipino, 15% W, 57% FRL, 14% ELL, 10% SPED. | SAT<br>HCPS II | +0.10<br>+0.22* | +0.16 |
| Bavarian et al. (2013) | CR | 6 years | 14 schools<br>1170 students (585E, 585C) | 3-8 | 48% AA, 27% H, > 50% FRL. | ISAT | | +0.17 |
| Promoting Alternative Thinking Strategies (PATHS) | | | | | | | | |
| MIE (2015) | CR | 2 years | 45 schools<br>2699 students (1,446E, 1,253C) | Year 5, 6 (U.S. grades 4, 5) | England. 30% FL, 22% ELL. | Key Stage 2<br>Year 5<br>Year 6 | <br>+0.03<br>-0.03 | 0.00 |
| Responsive Classroom | | | | | | | | |
| Rimm-Kaufman et al. (2007) | CQE | 1-3 years | 6 schools<br>1,401 students (769E, 632C) | 2-4 | 52% W, 22% AA, 21% H, 35% FRL. | CMT-Math | | +0.21 |
| Rimm-Kaufman et al. (2014) | CR | 3 years | 24 schools<br>2904 students (1,467E, 1,437C) | 3-5 | 41% W, 11% AA, 19% A, 24% H, 28% ELL. | SOL | | -0.13 |
| SSIS-CP | | | | | | | | |
| DiPerna et al. (2016) | CR | 12 weeks | 38 classes<br>402 students (210E, 192C) | 2 | 75% W, 17% AA. | STAR Math | | -0.03 |
| DiPerna et al. (2018) | CR | 12 weeks | 57 classes<br>696 students (341E, 355C) | 1 | 70% W, 24% AA, 9% H. | STAR Math | | +0.04 |

Table 8
*Mathematics Curricula*

| *Study* | *Design* | *Duration* | *k* | *Grade* | *Sample Characteristics* | *Posttest* | *ES by Subgroup/ Measure* | *Overall ES* |
|---|---|---|---|---|---|---|---|---|
| Early Learning in Mathematics | | | | | | | | |
| Clarke et al. (2015) | CR | 1 year | 129 classes<br>2116 students (1,134E, 982C) | K | OR, TX. 56% FRL, 38% ELL, 36% H, 8% SPED. | TEMA-3 | | +0.11 |
| enVisionMATH / Scott Foresman-Addison Wesley Elementary Math | | | | | | | | |
| Resendez & Manley (2005) | CR | 1 year | 35 teachers<br>645 students (352E, 293C) | 2, 4 | WA, WY, VA, KY<br>20% AA, 9% H, 10% ELL, 46% FRL. | TerraNova<br>Math Total<br>Computation | +0.10<br><br>-0.21 | -0.04 |
| Resendez & Azin (2006) | CR | 1 year | 39 classes<br>863 students (445E, 418C) | 3, 5 | OH, NJ<br>9% AA, 18% FRL. | TerraNova<br>Math Total<br>Computation | <br>-0.07<br>+0.05 | -0.01 |
| Resendez et al. (2009) | CR | 2 years | 44 teachers<br>659 students (349, 310C) | 2-3, 4-5 | MT, OH, NH, MA, KY, TN.<br>95%W, 19% FRL. | MAT<br>Conc. & Prob. Sol.<br>Math Computation<br>GRADE | <br>-0.13<br>+0.06<br>-0.06 | -0.04 |
| Strobel et al. (2017) | CR | 2 years | 33 teachers<br>495 students (285E, 210C) | 1-2, 4-5 | 24% W, 37% AA, 33% H, 15% ELL, 74% FRL. | TerraNova | | +0.02 |
| Everyday Mathematics | | | | | | | | |
| Vaden-Kiernan et al. (2015) | CR | 2 years | 48 schools<br>4467 students | K-5 | 51% AA, 73% FRL. | GMADE | | -0.01 |
| GO Math! | | | | | | | | |
| Eddy et al. (2014) | CR | 1 year | 79 teachers<br>9 schools<br>1,363 students (754E, 609C) | 1-3 | AZ, ID, IL, MI, OH, PA, UT, 36% AA, 35% H, 31% ELL, 35% FRL. | ITBS | | +0.01 |
| Investigations in Number, Data, and Space | | | | | | | | |
| Agodini et al. (2010) | CR | 1 year | 93 schools<br>4,019 students<br>(1,941E, 2,078C) | 1,2 | CT, FL, KY, MN, MS, MO, NY, NV, SC, TX.<br>23% AA, 32% H, 13% ELL. | ECLS-K<br>Grade 1<br>Grade 2 | <br>0.00<br>+0.09 | +0.04 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Gatti & Giordano (2008) | CR | 1 year | 77 classes<br>1363 students (729E, 634C) | 1,4 | AZ, MA, OR, SC<br>52% FRL, 27% H, 9% AA. | GMADE<br>Grade 1<br>Grade 4 | <br>-0.14<br>-0.31 | -0.23* |
| JUMP Math | | | | | | | | |
| Solomon et al. (2011) | CR | 5 months | 18 schools<br>267 students (163E, 104C) | 5 | Rural Canadian schools, Ontario. | WJ-III | | +0.23 |
| Math Connects | | | | | | | | |
| Jordan (2009) | CQE | 1 year | 139 teachers<br>1,897 students<br>(844E, 1,053C) | 2, 4 | 61% W, 14% AA, 16% H. | TerraNova<br>Grade 2<br>Grade 4 | <br>+0.08<br>-0.04 | +0.02 |
| Math Expressions | | | | | | | | |
| Agodini et al. (2010) | CR | 1 year | 90 schools<br>4,114 students<br>(2,036E, 2,078C) | 1,2 | CT, FL, KY, MN, MS, MO, NY, NV, SC, TX.<br>26% AA, 30% H, 10% ELL. | ECLS-K<br>Grade 1<br>Grade 2 | <br>+0.11*<br>+0.12* | +0.11* |
| Math in Focus | | | | | | | | |
| ERIA (2010) | CQE | 1 year | 678 students (125E, 553C) | 4 | NJ. 15% FRL, 30% minority, 12% SPED. | NJ ASK | | +0.25* |
| ERIA (2013) | CQE | 1 year | 33 classes<br>679 students (362E, 317C) | 3 | 59% minority, 58% FRL, 9% ELL. | ITBS | | +0.29 |
| Jaciw et al. (2016) | CR | 1 year | 18 teams<br>1,641 students (857E, 784C) | 3-5 | Clark County, NV.<br>47% H, 10% AA, 56% FRL, 11% SPED. | SAT-10<br>Problem solving<br>Procedures<br>Nevada CRT | <br>+0.12*<br>+0.14*<br>+0.05 | +0.10 |
| Saxon Math | | | | | | | | |
| Agodini et al. (2010) | CR | 1 year | 91 schools<br>4,083 students<br>(2,005E, 2,078C) | 1,2 | CT, FL, KY, MN, MS, MO, NY, NV, SC, TX.<br>21% AA, 40% H, 12% ELL. | ECLS-K<br>Grade 1<br>Grade 2 | <br>+0.07<br>+0.17* | +0.11 |

Table 9
*Benchmark Assessments*

| *Study* | *Design* | *Duration* | *k* | *Grade* | *Sample Characteristics* | *Posttest* | *ES by Subgroup/ Measure* | *Overal l ES* |
|---|---|---|---|---|---|---|---|---|
| Achievement Network (ANet) | | | | | | | | |
| West et al. (2016) | CR | 2 years | 89 schools 13233 students (6617E, 6616C) | 3-5 | MA, LA, IL. 87% AA, 15% ELL, 87% FRL. | State tests | | -0.09* |
| Acuity | | | | | | | | |
| Konstantopoulos et al. (2013) | CR | 1 year | 49 schools 11632 students (5816E, 5816C) | 3-6 | Rural, urban, and suburban schools in IN | ISTEP+ | | +0.19* |
| Konstantopoulos et al. (2016) | CR | 1 year | 55 schools 13944students (6972E, 6972C) | 3-6 | IN. 53% W, 27% AA, 12% H, 57% FRL 19% SPED. | ISTEP+ | | +0.13 |
| mClass | | | | | | | | |
| Konstantopoulos et al. (2016) | CR | 1 year | 55 schools 6249 students | K-2 | IN. 27%AA, 12% H, 57% FRL, 19% SPED. | TerraNova | | -0.22* |

Table 10

*Moderator analyses*

| *Moderator* | *k* | *ES* | *Q* | *p* |
|---|---|---|---|---|
| *Research design* | | | | |
| Randomized studies | 65 | +0.08 | 2.49 | *ns* |
| Quasi-experiments | 13 | +0.16 | | |
| *Grade level* | | | | |
| K-2 | 22 | +0.07 | 0.19 | *ns* |
| 3-6 | 40 | +0.05 | | |
| *Achievement level* | | | | |
| Low achievers | 8 | +0.08 | 0.89 | *ns* |
| Moderate and high achievers | 8 | +0.03 | | |
| *SES* | | | | |
| Low SES | 24 | +0.08 | 0.63 | *ns* |
| Moderate/high SES | 29 | +0.05 | | |
| *ELL* | | | | |
| High ELL | 7 | +0.09 | 0.73 | *ns* |
| Low ELL | 24 | +0.05 | | |

*Note*. *k* = total number of studies; *ES* = effect size; SES = socioeconomic status; ELL = English language learners

Table 11

*Programs Meeting ESSA Evidence Standards for Strong and Moderate Ratings*

| *Program* | *k* | *Average ES* | *ESSA Rating* |
|---|---|---|---|
| *One-to-One Tutoring by Teachers* | | | |
| Numbers Count | 1 | +0.33 | Strong |
| Math Recovery | 1 | +0.24 | Moderate |
| *One-to-One Tutoring by Paraprofessionals* | | | |
| Catch Up® Numeracy | 1 | +0.21 | Strong |
| Galaxy Math | 1 | +0.25 | Strong |
| Pirate Math | 1 | +0.37 | Strong |
| *One-to-Small-Group Tutoring by Teachers* | | | |
| Number Rockets | 1 | +0.34 | Strong |
| *One-to-One Tutoring by Paid Volunteers* | | | |
| Math Corps | 1 | +0.20 | Strong |
| *Small Group Tutoring by Paraprofessionals* | | | |
| FocusMATH | 1 | +0.24 | Strong |
| Fraction Face-Off! | 2 | +0.51 | Strong |
| ROOTS | 2 | +0.24 | Strong |
| *Programs Incorporating Technology* | | | |
| DreamBox Learning | 1 | +0.11 | Strong |
| Mathematics and Reasoning | 1 | +0.20 | Strong |
| SuccessMaker | 3 | +0.24 | Strong |
| *Instructional Process Programs* | | | |
| PAX Good Behavior Game | 1 | +0.32 | Moderate |
| *Whole-School Reform* | | | |
| Center for Data-Driven Reform in Education | 1 | +0.15 | Strong |
| *Social-Emotional Interventions* | | | |
| Positive Action | 2 | +0.16 | Strong |
| *Mathematics Curricula* | | | |
| Math Expressions | 1 | +0.11 | Strong |

| | | | |
|---|---|---|---|
| Math in Focus | 3 | +0.25 | Strong |
| | | | |
| *Benchmark Assessments* | | | |
| Acuity | 2 | +0.16 | Strong |

*Note*. *k* = total number of studies; *ES* = effect size.

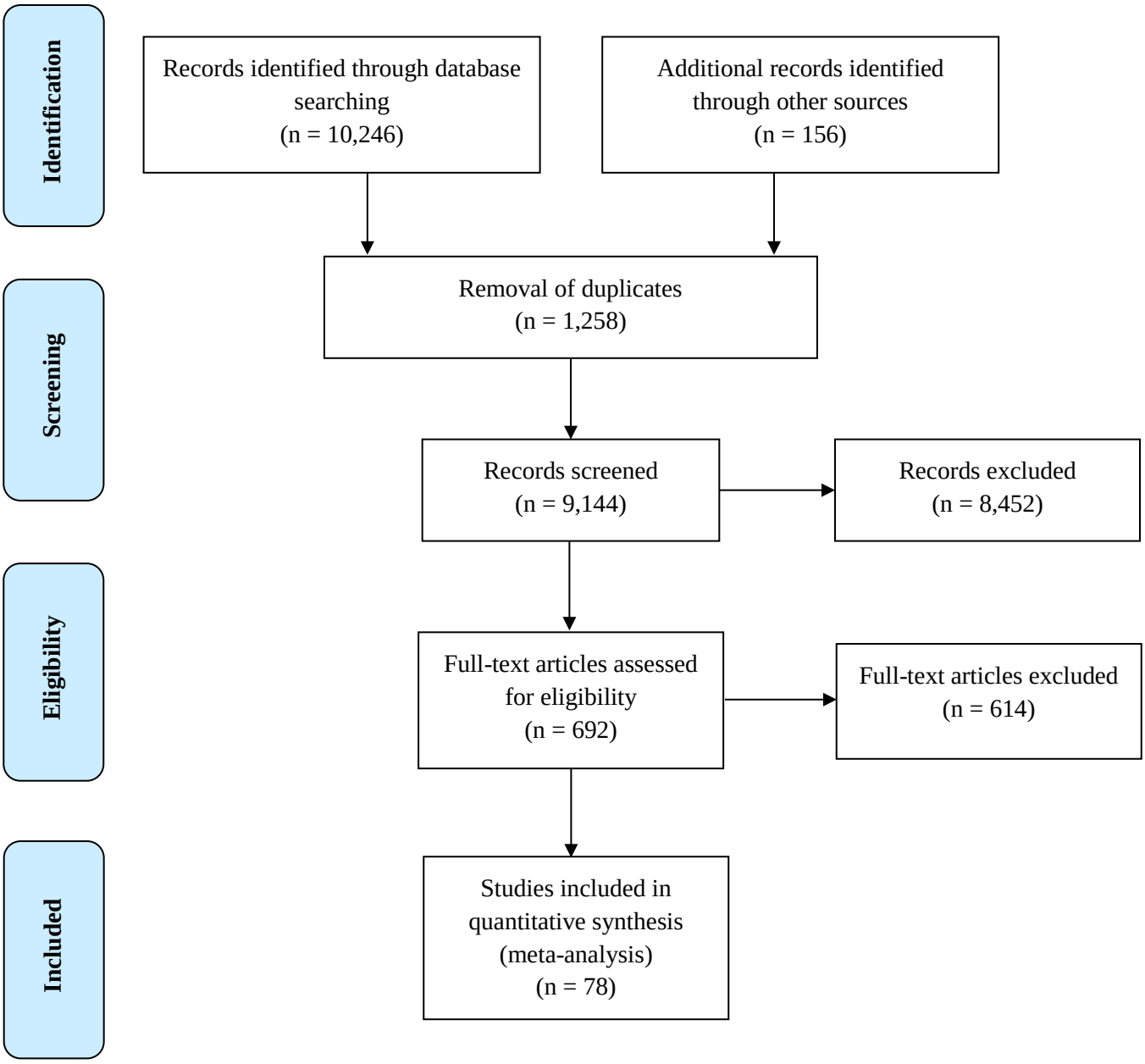


Figure 1. Selection procedures. Adapted from "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement", by D. Moher, A. Liberati, J. Tetzlaff and D.G. Altman, 2009, *PLoS Med*, *6*, p. 267.