Examining the Impact of a Tutoring Program Implemented with Community Support on Math Proficiency and Growth

Abstract

The current study evaluated the impact of a math tutoring program delivered in 20 schools to students in 4th through 8th grades by community members over one academic year. Students were randomly assigned to treatment and control groups. Multi-level linear and generalized linear mixed models were used to evaluate group differences in post-test scores and the probability of attaining the spring proficiency benchmark on two increasingly distal measures of math achievement. Intent-to-treat analyses identified higher achievement scores among students assigned to treatment on a measure of fact fluency and a computer adaptive measure of overall math achievement. Students assigned to treatment also had a higher probability of reaching grade-level benchmarks on the computer adaptive test. No statistically significant effects were observed on a state proficiency test. Implications for significant and null findings are discussed within the context of intervention content and delivery.

Keywords: mathematics intervention, tutoring, community-supports, RCT, tier 2

Examining the Impact of a Tutoring Program Implemented with Community Support on Math Proficiency and Growth

According to the most recent National Assessment of Educational Progress (NCES, 2019), only 41% of fourth graders and 34% of eighth graders in the United States are proficient in mathematics and performance declined for eighth grade students at or below the 25th percentile relative to previous years. These data suggest that many students experience difficulties with mathematics in school. As many as 17% of school-age children experience substantial mathematics difficulties, 7% of children will be diagnosed as having a mathematics learning disability, and an additional 5% to 10% experience persistent low achievement (Berch & Mazzocco, 2007; Geary et al., 2007; Shalev et al., 2005). Collectively, students that struggle with mathematics in school face several negative outcomes. These students are not as likely as their peers to succeed in or even complete high school and they lack the quantitative literacy necessary for a variety of careers (Claessens & Engel, 2013; Parsons & Bynner, 1997). Unfortunately, mathematics difficulties begin as early as pre-school and, without intervention, persist through intermediate grades (e.g., Duncan et al., 2007; Morgan et al., 2011). Such realities for math education in the United States illustrate the need to provide students who are at-risk for mathematics failure with supplemental school-based supports.

Evidence-based Math Intervention

Schools are increasingly turning to educational frameworks that focus on prevention and early intervention (Jimerson et al., 2016), because such frameworks represent a promising and systematic approach to address the needs of more students with academic difficulties sooner than traditional frameworks of service-delivery. Often referred to as response to intervention (RtI) or multi-tiered systems of support, such frameworks are comprehensive approaches to improving educational systems. They prescribe high-quality, research-based core instructional practices for all students; intervention and progress monitoring for students who are low-achieving; and individualized, high-intensity support for students requiring special education (Burns et al., 2016). Data are used within these frameworks to inform decisions and supplemental interventions are provided to students who demonstrate risk for academic failure based on screening measures or minimal progress in response to core instruction (Gersten et al., 2009). Although every state has an initiative to promote and expand these frameworks (Jimerson et al., 2016), only about 59% of elementary schools and 48% of middle schools report implementation in math (Spectrum K-12, 2010).

One of the most significant implementation barriers for such frameworks is a lack of intervention knowledge and resources, including perceptions of intervention compatibility, time demands, and material needs (Long et al., 2016; Spectrum K-12 et al., 2010). Intervention delivery for prevention and early intervention frameworks has typically followed either an individualized problem-solving approach or a standard protocol approach, with standard protocol approaches having advantages, both logistically and evidentiary, for at-risk students not yet identified for special education (Fuchs & Vaughn, 2012). For math, there are a limited number of evidence-based standard protocol interventions designed to target foundational skills for students beyond third grade. According to *What Works Clearinghouse* (WWC, 2016), when filtering by grade and either supplemental or small group intervention, there are only three math interventions that have received a rating of potentially positive or positive for students in grades four through eight (Cognitive Tutor® Algebra 1, Fraction Face-Off!, and Odyssey® Math). Other evidence summaries report a similar paucity of math interventions. The *National Center on Intensive Intervention* includes 10 individual or small group intervention programs or strategies directed toward the elementary or middle school levels as offering convincing or partially convincing evidence of effectiveness. Further, although there are many peer-refereed articles published on mathematics intervention outcomes, in 51.5% of these studies researchers, rather than school professionals, were the implementers, suggesting that these interventions might be in the development phase and therefore not widely available for practitioner use (DeFouw et al., 2018). Finally, the majority of these intervention studies addressed whole number knowledge and skills appropriate for students within the early elementary school grades. Interventions with strong evidence for improving rational number skills for older students are comparatively rare.

Even when evidence-based standard protocol interventions are available, teachers report that finding time and resources to integrate the intervention within existing routines is a substantial barrier, and almost 89% indicated a need for additional implementation support (Long et al., 2016). These data correspond with other surveys suggesting such barriers can be powerful constraints when it comes to providing students' academic supports outside of core instruction (e.g., Bambara et al., 2009; Bosworth et al., 1999; Durlak & Dupre, 2008). Unfortunately, the persistence of such barriers impacts the integrity with which prevention and early intervention frameworks are successful. In one national evaluation of RtI, nearly 40% of schools neglected to administer intervention supports as intended resulting in an overall failure to observe positive effects of the framework (Balu et al., 2015; Fuchs & Fuchs, 2017).

Encouragingly, recent research shows that K-5 validated math intervention programs delivered by *either* teachers or paraprofessionals produced strong effects, particularly for low achieving students and students from low income backgrounds (Pelligrini et al., 2018). Therefore, licensed professionals may not be required to deliver evidence-based intervention programs in order to produce positive outcomes for students. The status of the research base suggests more research is necessary to identify math interventions with evidence for effectiveness in typical school settings, as opposed to efficacy evidence produced with strong involvement of research teams, particularly for intermediate elementary and middle school students. Further, given the high numbers of students that need additional mathematics supports and the limited time and support presently available in schools to provide students with evidence-based interventions, it also appears necessary to identify feasible delivery mechanisms for such interventions. Community-based tutoring might be a reasonable alternative.

Community-supported Tutoring

The idea of using community-based volunteers or non-educators to provide substantive academic supports in schools is neither novel nor without evidence. Most schools have access to multiple community- or business-based volunteers, and these volunteers can make a significant positive impact on student learning outcomes (Ritter et al., 2009; Slavin et al., 2011). Moreover, some community organizations offer fully-developed interventions that align with educational best practices and closely resemble standard protocol interventions (Jacob et al., 2016; Markovitz et al., 2014). The potential for such programs is particularly promising in the context of the logistical and knowledge barriers many schools face when attempting to deliver educational interventions. Early intervention is often a paradigm-shift in educational settings that rely on conventional “wait-to-fail” models for educational service delivery (Fuchs & Fuchs, 2006). The experience and knowledge required for effective resource reallocation can be a considerable challenge (Noell & Gansle, 2016). Moreover, few practitioners appear adequately familiar with the concepts and practices necessary for successful intervention within prevention and early intervention frameworks (Vujnovic et al., 2014). These knowledge barriers are exacerbated by basic resource demands for providing evidence-based interventions, which in math require schools to provide each struggling student up to 90 minutes of intervention support per week (Gersten et al., 2009).

Community-based organizations help address these barriers by providing schools with resources, time, and expertise. Such organizations also have the advantage of being able to specialize their topic area. In math, this translates to knowledge about evidence-based math interventions and technically adequate and instructionally informative data-based decision-making practices, as well as being able to generate resources necessary for implementation. Community-based organizations that develop specialized expertise in an educational area can form inter-organizational partnerships with schools that facilitate effective implementation of math interventions (Aarons et al., 2011). In other academic domains, such as reading, inter-organizational partnerships using community support have also proven to be cost-effective ways to deliver intervention supports within the school context (Hollands et al., 2016).

Math Corps

The ecology of student development includes macro-level structures that influence student outcomes (Bronfrenner, 1987). Within this broader ecology, public policy is inherently designed to influence issues of the broader public interest, and the federal AmeriCorps program is an example of a policy initiative that translates to increased local community-based resources for implementing educational interventions. The AmeriCorps program helps generate nearly seven billion hours of community service annually (Corporation for National and Community Service, 2018), and a portion of that time has been allocated to educational programs that exist to deliver standard protocol interventions for struggling students. The Math Corps program is an intervention program for students in grades four through eight that uses AmeriCorps members to deliver math intervention support to struggling students. It provides schools with materials for implementing math interventions, recruits and trains tutors and school staff, and engages in ongoing coaching (see Method section for details).

Previously, a wait-list randomized control trial of the Math Corps program was conducted with 550 students in grades four through eight who were at-risk for mathematics failure over the course of one semester (Authors, 2019). Results on a computer-adaptive, comprehensive measure of mathematics performance indicated that students receiving Math Corps tutoring outperformed the control group, yielding a small positive effect size (d = .17). Effect sizes were larger (d = .24) when the designated optimal dose of the intervention was achieved (i.e., 12 weeks of intervention with an average of 60 min/week). This study represents initial research in the area of mathematics regarding the promise of community-supported delivery of Tier 2 interventions. However, the prior study did not examine outcomes on other measures of interest, such as math skills more proximal to the intervention focus or distal outcomes of policy importance like state math tests. Given that mathematics is hierarchical and builds upon core foundational skills, analysis of the extent to which other math skills improve would be useful. Previous research has indicated that the amount of exposure to and the number of opportunities to practice any given mathematics subskill can be substantial before mastery occurs across both basic (Burns et al., 2015; Stickney et al., 2012) and complex (Nelson et al., 2018) tasks. In particular whole number fact fluency appears to be a central underpinning of overall mathematics proficiency, permitting greater access to higher level content including more complex whole and rational number operations and understanding (Bailey et al., 2014; Jordan et al., 2017). Such considerations reflect increasingly important aspects of replication studies for intervention research (Travers et al., 2016).

Purpose

The purpose of the present study was to evaluate the impact of Math Corps as a supplemental math intervention delivered by AmeriCorps members across the 2017-2018 academic year. The study was guided by the overall goal of evaluating the extent to which end-of-year math achievement differed for students randomly identified to be served by the program or not served by the program. There were three primary math outcomes of interest arranged from most proximal to most distal based on the program's theory of change: basic fact fluency, broad-based math achievement as measured by a computer adaptive test, and end-of-year performance on the state test. The following research questions framed the study: (1) to what extent do students assigned to Math Corps demonstrate higher test scores on a basic fact fluency measure relative to students assigned to a control group? (2) to what extent do students assigned to Math Corps demonstrate higher posttest scores on a computer adaptive measure of math achievement relative to students assigned to a control group? (3) to what extent do students assigned to Math Corps demonstrate higher posttest scores on the end-of-year state achievement test relative to students assigned to a control group? In addition to addressing the above research questions by comparing the performance of students in both groups on each measure, we also compared the performance of students in both groups to examine the degree to which students met criterion-based levels of proficiency for the computer-adaptive and year-end state tests using logistic regression. The value of this approach is often advocated by researchers as screening and state test scores are interpreted as such in practice (Fuchs & Fuchs, 2017). That is, in addition to examining students' overall scores, educators routinely examine whether students are above or below a meaningful grade-level benchmark. The latter form of interpretation is particularly relevant as attainment interpretations typically guide resource allocation decisions within a tiered support model.

Method

Participants

Twenty schools across the state of Minnesota participated, including six middle schools serving students in grades 6-8, one intermediate school serving grades 3-5, and 13 elementary schools serving grades K-5. Nine schools in the sample were suburban (45%), five were urban (25%), and six were rural (30%). All 105 schools that received Math Corps services in the 2017-2018 school year were invited to participate, resulting in the 20 schools that volunteered. In addition, the project team evaluated historical program data for potential sites to ensure schools served enough low achieving students (at least 50 for a full-time tutor) to support a treatment and a control group.

All students meeting eligibility criteria within those schools were included in the randomization process. Students who are eligible for Math Corps must (1) score below proficiency on the state test from the previous year (the Minnesota Comprehensive Assessment [MCA]) and (2) score below a grade-level benchmark on the fall administration of STAR Math (Renaissance Learning, 2018).

In the current study, a total of 924 students were identified as eligible for Math Corps at the 20 participating schools. In most cases, only one tutor was placed by the program at each school; however, four of the 20 schools received support from more than one tutor. Of the 924 students eligible for program support, 750 were randomly assigned to treatment (n = 484) and control (n = 266) groups. An additional 174 students were assigned to a waitlist group to ensure interventionist caseloads would remain full if treatment students moved or withdrew consent from participating. Data for waitlist students were not retained for the current study. Following randomization, 37 students (34 from treatment and three from control) withdrew consent and four students (three from treatment and one from control) were determined ineligible due to receipt of special education math services. Table 1 provides student demographic data for the students in the Math Corps and control groups. Chi-square analyses revealed no statistically significant differences between-groups on grade, gender, and ethnicity. Participant cohorts by school ranged from 15 to 85 students (M = 35, SD = 19). Recruitment yielded an appropriate sample size.

There were slightly more female students (51.3%) relative to male students (48.6%) in this sample. Across both groups, a majority of students were White (53.4% and 52.9%) followed by Black (24.5% and 23.3%), Latinx (10.6% and 12%), Asian-American (8.9% and 6.9%), Native North American (1.2% and 0.5%), and Other (0.7% and 2.3%). The sample student distribution across elementary and middle schools was generally commensurate with the overall distribution of students served by Math Corps. For example, across all racial categories, the largest difference between the study sample and broader population served by Math Corps was approximately 8% (53% of students in the sample were White relative to 46% of students within the full Math Corps population).

Intervention Overview

The instructional focus of Math Corps is on improving whole and rational number understanding. Math Corps uses intervention strategies that target specific skills (e.g., adding fractions with like and unlike denominators) required to effectively work with whole and rational numbers. In the present study, Math Corps was implemented by full-time AmeriCorps members who operated as tutors embedded within each school. Math Corps tutors had no preexisting professional preparation in standard protocol interventions for math. Each tutor attended a three-day training session in late summer, two additional days of training in the fall (October and

November), and received monthly coaching sessions from a school-based (internal) coach and a program (master) coach, both of whom were fully-trained in the Math Corps program model.

There was a total of 26 tutors among the 20 participating schools. Most schools had one tutor (*n* = 16) but two schools had two tutors and two schools had three tutors. All tutors were hired and trained by the Math Corps organization. Tutors were observed monthly between September and March or April by internal and/or master coaches, resulting in six observations by the internal coaches and five from the master coaches. In September tutors were observed administering the STAR assessment but during the remaining observations tutors were observed delivering specified lessons. During all observations, coaches used a standardized observation form in which implementation fidelity was assessed using a checklist with 19 items aligned with various components of the intervention. In addition, coaches assigned an engagement quality rating and instructional delivery quality rating to tutors after each observation. Both ratings were scaled from one to five with five representing the highest quality rating. In the present study, tutors tended to implement the intervention as intended with average adherence equal to 91%, delivery quality equal to 4.32 (*SD* = 0.75), and engagement quality equal to 4.44 (*SD* = 0.56).

Math Corps content. Math Corps delivers intervention in the form of instructional lessons, which varied across grade levels. In fourth grade, 22 lessons were delivered across approximately 17 weeks, in 5th grade 25 lessons were delivered across 21 weeks, in 6th grade 31 lessons were delivered across 21 weeks, in 7th grade 27 lessons were delivered across 22 weeks, and in 8th grade 20 lessons were delivered across 18 week. Lessons used one of three intervention components to improve targeted subskills required to work effectively with whole and rational numbers. The first component included conceptual-based instruction using the Concrete, Representational, Abstract (CRA) approach (e.g., Witzel et al., 2003). The second component focused on procedural accuracy and included direct instruction followed by supervised practice with Cover, Copy, and Compare (CCC; Skinner, Turco, Beatty, & Rasavage, 1989). The third component used Cognitive Strategy Instruction (CSI) to support development of the skill for word problem solving (Montague, 1997). Intervention components were applied in a sequence for each skill. For example, students first received CRA to better develop the conceptual basis for adding fractions with dissimilar denominators; then received CCC to become efficient at accurately applying the corresponding computational strategies; and then received CSI to be able to solve word problems involving fractions with unlike denominators. Tutors used a brief informal assessment connected to a given intervention to guide student progress from one intervention to the next. Tutors used 85% correct on that “stop and check” as a general rule before advancing to the next intervention components. Tutors also delivered short duration fact fluency practice using Explicit Timing (e.g. Van Houten & Thompson, 1976) at the end of intervention sessions.

All intervention support was provided during the school day to groups of two students, consistent with standard Math Corps program procedures. Student pairs were determined by performance on a brief curricular inventory, which samples two to five items from each unit, as well as the scheduling constraints of the school. Tutors worked with school staff to provide student pairs either three 30-min or two 45-min intervention sessions per week. Across the academic year, and accounting for short weeks and interventionist or student absences, the average min per week for students was equal to 69.33 (*SD* = 18.93) over an average of 21.89 weeks (*SD* = 10.34). Participating students completed between 10.11 (eighth grade) and 13.06 (sixth grade) lessons across the school year, indicating that many students required extensive practice with early math concepts. Dosage information relates only to students that were assigned to receive Math Corps and actually received the intervention. An additional 59 students did not have any treatment time recorded, but were included in analyses.

Control. Students assigned to the control group in the present study did not receive Math Corps interventions but were allowed to receive other school-based services. Surveys were sent to school staff two times during the study to determine the frequency and type of other mathematics services throughout the school year. Approximately 43% of control group students received more than 30 minutes per week of supplemental support for at least one month during the year. The nature of additional support varied between sites, but most frequently took the form of supplemental small-group pre-teaching/re-teaching activities with a teacher or teaching assistant; no evidence-based practices were reported.

Randomization

After eligible students were identified, block randomization procedures were used to place students in either treatment, control, or waitlist conditions (Imbens & Rubin, 2015). All eligible students identified by each tutor within a school were split into two blocks by the median STAR score. Within both blocks, randomization was conducted via an R script using uniform variables 0-1. These numbers were then randomly attached to each student and then lists were ordered from least to greatest based on the random number generated by R. Students were assigned to groups based on the list order. For a full-time tutor, the first 12 students in each block were assigned to the treatment group, the next 4 were assigned to the waitlist, and the remaining students were put into the control group. For a half-time tutor, the first 6 students in each block were assigned to the treatment group, the next 2 students assigned to the waitlist, and the remaining students placed in the control group. The overall ratio of the sample was designed for a 3:2 distribution of students between treatment and control groups.

A series of analyses were conducted on the resulting sample to establish baseline equivalence. Chi-square analyses revealed no statistically significant differences between groups on the distribution of grade, gender, or ethnicity. Further, a paired *t*-test examining state test scores from the previous year revealed no significant differences between groups. However, tests for baseline equivalence conducted on STAR Math pretest scores revealed a significant difference between groups (t (693) = 2.79, p = .005). The mean fall STAR Math score for students in the treatment group was equal to 646.77 (*SD* = 102.91) and the mean score for students in the control group was equal to 667.70 (*SD* = 83.79). An analysis of the distribution of fall STAR Math scores revealed a disproportionate number of students in the treatment group with scores far below the mean in each grade level.

After examining potential reasons for this difference, it was determined that although assignment to treatment condition was random, the randomization process resulted in disproportionate assignment across conditions in which more students with very low STAR Math scores were assigned to the treatment condition. This occurred because the evaluation team established two blocks of students by median score within a school, rank-ordered the scores lowest-to-greatest within block, assigned random numbers to each case within the block, re-ordered lowest-to-greatest, and then made condition assignments starting with treatment. For this process, a single fixed ‘seed order’ of random numbers was assigned across all blocks that—by chance—had more lower values earlier in the seed order, which meant more students with low STAR Math scores were assigned to treatment. Several potential solutions were explored to address this issue, including re-randomization, propensity score weighting, and trimming potential outliers from the data file. Inferential results did not vary across methods for adjusting for baseline differences; however, the most reasonable approach included trimming potential outliers from the data file, identified as STAR pre-test scores outside of two standard deviations from the mean (Kutner et al., 2004). This resulted in the exclusion of 31 students originally assigned to the treatment group and one student originally assigned to the control group. After trimming the data file for potential outliers, there were no significant differences in baseline STAR scores between groups. The resulting file for analysis included 416 students assigned to Math Corps and 259 students assigned to the comparison group (N = 675).

Measures

Minnesota Comprehensive Assessment (MCA). Students completed the MCA in math at the end of the academic year. The MCAs are the statewide accountability tests used by the state of Minnesota. Beginning in the third grade and extending to eighth grade, students in Minnesota complete the MCA in math. The MCA is a standards-based assessment—the items on the test are constructed to align with the curricular standards for the state of Minnesota. Thus, students' scores on the MCA are interpreted as the degree to which students have mastered grade level content. MCA scores for math and reading range from 0-100, with scores of 50 and above signifying proficiency. The MCA-III has adequate evidence for internal consistency (r = .78 to .95; Minnesota Department of Education, 2017). Because the MCA is a standards-based assessment, validity evidence is primarily derived from content-related evidence (e.g., test blueprint aligned with standards, expert item writers) and construct-related evidence (e.g., inter-scale correlations, high functioning items). In the present study, MCA testing was conducted by trained test officials at school sites in accordance with state guidelines.

STAR Math. Students completed STAR Math at the beginning and end of the academic year. STAR Math is a computer-adaptive test (CAT) with scaled scores ranging from 0-1400. The vendor reported split-half reliability estimate for grades 1-12 is equal to .94 (Renaissance

Learning, 2018). In a recent evaluation, concurrent validity correlations between STAR Math and the Minnesota state test were equal to .74 (grade three) and .73 (grade five). Predictive validity coefficients (fall to spring) were equal to .60 (grade three) and .75 (grade five). The publisher for STAR Math provides spring benchmarks for performance that were derived using diagnostic accuracy analyses using MCA proficiency as the criterion. Those benchmarks are used to identify students as on-track for proficiency or below proficiency. In the present study, we used those benchmarks as the metric of interest in the proficiency analyses. Pre-test STAR Math data were obtained by tutors, prior to randomization, in accordance with program guidelines for screening—those data are used to identify students for Math Corps support. Post-test STAR Math data were obtained by research assistants who were blind to group assignment; with one exception. One school district regularly used STAR Math as part of their universal screening procedures and therefore collected these data outside of the Math Corps procedures. In that case, school officials oversaw post-test data collection using their standard school-wide procedures.

Subskill Mastery Measures. Subskill mastery measures are a single-skill version of curriculum-based assessment that were used to assess fact fluency (Shapiro, 2011). Students completed a timed (1-min) test of fact fluency in the fall (administered by Math Corps tutors) and spring (administered by research assistants blind to group assignment) of the academic year. Parallel eighty-item mixed computation measures (Foegen, 2000; Foegen & Deno, 2001) were used to assess fluency with basic facts. The subskill mastery probes required students to solve single-digit combinations (0-9) in addition, subtraction, multiplication, and division. All probes were scored as the number of problems correct in one min and were identical across grades. Math Corps uses a fluency proficiency goal of 30 problems correct in one min, which approximates the rate at which previous research (e.g., Burns, 2005) using the subskill mastery measures provide sufficient reliability evidence (internal consistency r = .91-.92; test-retest r = .85; alternate forms r = .82). In the present study, the correlation between students' fall subskill mastery scores and STAR Math was .57, which is generally consistent with studies examining the relationship between basic computation skills and broad measures of achievement (Foegen, 2000; Jitendra et al., 2005; Thurber et al., 2002). In the present study, fact fluency data were obtained immediately before STAR Math testing.

Analysis Procedures

Power analysis. An a priori power analysis was conducted using a conservative estimate of effect (d = .20) based on Author et al. (2019) and similar intervention studies, with consideration of a 60:40 distribution of students in the experimental and control groups, an alpha level of .05, and the inclusion of the planned covariates for pretest, grade, gender, and race. The resulting power estimate was equal to .80 for a sample size of 720.

Missing data. Missing data differed by posttest outcome. The largest instances of missing data were observed for pre (12%) and post (13%) fact fluency tests. Approximately 10% of cases were missing MCA data. Lower levels of missing data were observed for STAR Math at pre-test (2%) and post-test (8%). A series of chi-squared tests indicated that there were no differences in missing data across group assignment for STAR Math or MCA data; however, students assigned to the control group were significantly more likely to have missing fact fluency data at post-test (22%) relative to students assigned to receive Math Corps (7%). This was largely due to a small number of sites that opted to oversee STAR Math testing, but failed to uniformly administer the fact fluency test. Regardless it is not likely that data were missing completely at random (MCAR) and instead were missing based on factors known in the data set

(MAR). To account for missing data in the analytic models we used multiple imputation to simulate 40 additional data sets with estimated values for missing data (Graham et al., 2007; Rubin, 2004). The multiple imputation procedure was run using SPSS (v. 25) and included multivariate normal regression paired with an iterative Markov chain Monte Carlo simulation process. To facilitate accurate estimation, imputation models included a series of predictors including STAR Math data, MCA data, fact fluency data, gender, ethnicity, grade level, site, and group assignment. Estimates from each simulated data set were pooled to provide a single estimate. Results from analytic models using the pooled estimates were compared to identical models in which listwise deletion was used. No differences were observed in the statistical significance or magnitude of effects between models that used imputed data and those that used listwise deletion.

Analytic models. As discussed, all three outcomes of interest in the present study were evaluated on a continuous scale and two of the three outcomes were also evaluated on a dichotomous (i.e., benchmark attainment) scale as educators routinely make decisions using both interpretations. In both cases, there was clustering inherent in the data. More specifically, students were nested within schools and tutors. However, given the near 1:1 correspondence between tutors and schools, fitting a three-level model or partially nested model was not possible. Given that tutors are likely the most meaningful cluster, we fit multi-level linear regression models to assess continuous outcomes and generalized linear mixed models to assess benchmark attainment, with students clustered by tutor. When examining benchmark attainment, previously established cut scores associated with the state test and STAR Math were used dichotomous students' spring math achievement. In the case of the MCA, a score of 50 or greater is associated with proficiency (Minnesota Department of Education, 2017). In the case of STAR

Math, diagnostic accuracy analyses completed by the publisher provide specific grade-level benchmark scores associated with proficiency (Renaissance Learning, 2018). The use of generalized linear mixed models permitted an evaluation of potential differences in the degree to which students met those criteria across groups. In addition to a dichotomous variable indicating treatment assignment, each model included students' pre-intervention scores and dichotomous variables for gender and race. It is important to note that due to low cell sizes for some race categories, students identifying as American Indian, Multi-Race, Latino, and Pacific Islander were collapsed into a single "other" category for modeling purposes—the association between race and math achievement was not a focus point of the study. All analyses followed an intent-to-treat framework in which all students were included in the analysis according to their original group assignment regardless of their experiences during the school year. Significance values in all models were adjusted for multiple comparisons (N = 5), resulting in a significance level of $p < .01$.

Results

Descriptive

Means and standard deviations for outcomes across time and groups are displayed in Table 2. On average, students in both groups tended to answer between 16 and 17 problems correct on the timed fact fluency pre-test. All students increased their fact fluency scores across the academic year. The average number of problems answered correctly at post-test among students assigned to receive Math Corps and those in the control group were equal to 23.29 and 19.46, respectively.

As discussed, students assigned to Math Corps had slightly lower STAR Math scores (approximately 10 points) at pre-test relative to students in the control group. At post-test, the average score for students in Math Corps was 11.22 scaled score points higher than the average score for students in the control group. In addition, a greater percentage of students assigned to Math Corps met the end-of-year benchmark for STAR Math (27%) relative to students in the control group (15%). Students assigned to Math Corps and student in the comparison group had similar average scores on the MCA (42.19 and 41.47, respectively). 19% of students receiving Math Corps passed the state test compared to 15% of students in the comparison group.

Multi-Level Linear Regression Models

The first set of research questions was related to student outcomes on a continuous scale. That is, to what degree did student scores on the fact fluency test, STAR Math test, and state test differ as a function of group assignment? Results from the three multi-level linear regression models of interest are displayed in Table 3. All models controlled for students' prior achievement (mean centered), grade level, gender, race. The referent for grade was always fourth grade and the referent for race was always White.

In regard to fact fluency, a statistically significant and positive association with post-test fact fluency scores were observed for pre-test fact fluency scores (B = 0.79). Assignment to Math Corps was associated with a statistically significant and positive effect on post-test fact fluency scores. Specifically, assignment to Math Corps was associated with a 3.48 increase in the number of problems correct on the fact fluency post-test relative to students assigned to the control group.

When examining STAR Math scores, statistically significant and positive effects were observed for fifth through eighth grade. That is, students in later grades had significantly larger scores, which is generally consistent with the scaling of STAR Math. In addition, we observed a positive and statistically significant effect among boys (B = 12.81). Assignment to Math Corps was associated with a statistically significant and positive impact on post-test scores equivalent to an advantage 16.63 scaled score points. In regard to students' state test scores, Black students were associated with lower state test scores relative to White students (B = -3.10). Unlike the models examining fact fluency and STAR Math, group assignment was not a significant predictor of students' MCA scores.

Generalized Linear Mixed Models

Two of the three outcomes of interest were also examined in dichotomous terms—students either met (1) or did not meet (0) a predefined criterion of proficiency on each outcome measure. Results from the two final GLMMs of interest are displayed in Table 4. All effects in Table 4 are expressed as the log-odds of attaining proficiency on the measure of interest. Throughout the results, we convert log-odds—the primary output of interest in GLMMs—to probabilities.

In regard to STAR Math, the log odds of meeting the grade-level spring benchmark significantly increased for every scaled score unit increase above the fall grade-level mean. The predicted probability of meeting the end-of-year STAR Math benchmark was higher among students assigned to Math Corps (p < .01). For example, among fourth grade students assigned to Math Corps, the predicted probability of success was .29 compared to a predicted probability of success equal to .14 among those in the comparison group.

When examining MCA proficiency (i.e., scores at or above 50), an increase of one fall STAR Math scaled score above the grade-level mean was associated with a small increase in the log-odds of meeting MCA proficiency (.01). Further, students in fifth grade had a lower log odds of meeting the MCA benchmark relative to fourth grade students. Black students also had a lower log odds of success relative to White students. Similar to the linear regression results, the impact of group assignment on the probability of meeting the MCA benchmark was non-significant.

Discussion

In the present study we evaluated the impact of Math Corps, a community-supported supplemental math intervention program, on three outcomes aligned with the program’s instructional focus. All analyses adopted an intent-to-treat approach in which students were included in the analysis regardless of their exposure to Math Corps. Statistically significant and positive results were observed for Math Corps’ impact on the two more proximal outcomes of interest—fact fluency performance and STAR Math performance. Yet state test scores for students assigned to receive Math Corps were not statistically different from those of students not assigned to Math Corps. These effects are discussed in detail below.

Math Corps and Student Math Achievement

Fact Fluency. Of the three outcomes considered in the present study, fact fluency performance was most closely aligned, or proximal to, Math Corps support. That is, tutors were trained to deliver 3-5 min of explicit fact fluency support as part of each session. In the present study, assignment to Math Corps was associated with an unstandardized effect size of 3.48 problems. The effect of Math Corps on fact fluency scores is particularly notable when contextualized by the average growth rate of control students across the full year—3.09 problems. In other words, the effect size observed for Math Corps was larger than the average annual growth of a typical student. Yet, it is also worth noting that an overwhelmingly large proportion of students across groups failed to correctly answer 30 or more fact fluency problems within one minute. This is relevant because automatic retrieval of basic math facts is generally considered to hold value for more complex math operations as well as a variety of tasks outside of school (Fuchs et al., 2014; Nelson et al., 2016). The relatively low levels of automaticity among the students included in the present study generally conform to findings that students who struggle on grade-level math assessments may also struggle with foundational skills (Jordan et al. 2013; Vukovic et al., 2014), and underscores the need to consider support for fact fluency skills beyond the grade-levels that fact fluency is most closely aligned with the core instruction curriculum (e.g., second and third grade).

STAR Math. STAR Math covers a full range of grade-level standards, many of which fall outside the scope of the Math Corps intervention which focuses on whole and rational number understanding. The underlying assumption of Math Corps is that advancing whole and rational number skills will confer benefits to the those skills, but also other skills that rely on whole and rational number understanding (e.g., geometry; Gersten et al., 2009; NMAP, 2008). Although STAR Math assesses a broad range of content and is therefore a somewhat distal outcome, the test is vertically scaled and grade independent, making it a useful broad-based measure of student growth.

In the present study, the impact of assignment to Math Corps was 16.63 scaled score points. The expected weekly growth among students in the sample (based on initial scaled score and grade level) ranged from .80 to 3.30 with an average of 1.76 (SD = .60). Thus, one interpretation of the unstandardized effect of group assignment is to consider the amount of additional growth added by the intervention. By this interpretation, the average student who received Math Corps demonstrated post-test scores that were the equivalent of 4.85 to 20 weeks (M = 9.09 weeks) of additional math growth compared to the average student who did not receive Math Corps. The impact on STAR Math scores also translated into an increased probability of meeting the end-of-year benchmark among students receiving Math Corps support.

MCA. Unlike the results observed for fact fluency and STAR Math, the impact on students' state test scores—including the rate at which students met proficiency criteria—was not statistically significant. Performance on the state test was the most distal outcome of interest in the present study. The Minnesota state test assesses a limited range of content—items are created to align with specific grade-level standards and do not explicitly address content that falls above or below those standards. Thus, improvements in understanding of precursor concepts (e.g., fraction arithmetic among eighth grade students) may be assumed and not directly tested. The nature of state test construction is relevant because many students who qualify for supplemental support are far behind grade-level expectations. For those students, relative to computer adaptive tests such as STAR Math, the state test may not be sensitive to gains in math skills outside the grade-level standards. As such, the state test is not designed to capture student growth, but proficiency on grade-level standards at the conclusion of the academic year (Shapiro & Gebhardt, 2012). In the present study, Math Corps resulted in significant improvement in math achievement as measured by STAR Math, but not the state test. Thus, although students improved their overall math achievement as a result of Math Corps, that improvement was not sufficient to produce differences on the state test. Intervention studies do not often evaluate impact on distal standardized measures, and thus the current findings are generally consistent with other findings that greater effects are typically observed for measures more closely aligned to treatment (DeFouw et al., 2018; Gersten et al., 2009; Jitendra et al., 2018).

As discussed, Math Corps is composed of a series of lessons for each grade level. Those lessons were created to align with state standards in the domain of numbers and operations, but they were not tied to a single grade level so as to permit students to receive needed intervention on precursor skills. For example, the current sixth grade lesson sequence includes 31 lessons. By design only 10 of those lessons are directly aligned with sixth grade content, 9 of which are the last lessons in the entire sequence. Many students who qualify for Math Corps have deficits on skills that precede grade level content, and math learning is generally characterized by progressive skill development that assumes precursor skill levels are established for more complex levels to be obtained (e.g., NMAP, 2008; Nelson et al., 2018). Because Math Corps provides support for precursor skills and has a mastery learning orientation, many students did not receive intervention on grade-level skills. For example, data from the program showed that only 10% of sixth grade students reached grade-level content, with an average of 11.34 (*SD* = 6.04) lessons completed across grades.

Implications for Community-Supported Educational Intervention

Overall, these findings add to the literature for community-supported math interventions in several ways. First, results support the potential for inter-organizational partnerships to implement evidence-based interventions in math (Aarons et al., 2011), and thus are relevant to broader calls for school psychology to better understand how community and school systems work together when implementing evidence-based interventions (Kratochwill, 2007). Systems designed to deliver social services (i.e., schools) likely need to be understood as distinct from systems designed to support implementation of innovative and effective practice (Wandersman et al., 2008). In the current study, Math Corps provided all training, data collection systems, and materials, but the school system played an integral role in identifying the community-based AmeriCorps member to serve as an interventionist, acclimating them to the local school environment, and ensuring they received sufficient coaching support for success. The mutual contribution of resources occurred across common phases of implementation and serves as a model for how inter-organizational partnerships could feasibly implement evidence-based practices in public service settings (Aarons et al., 2011).

These findings also extend the overall empirical support for service-based educational interventions that employ tutoring. When service-based interventions utilize tutoring, they employ a form of educational instruction for low achieving students that has increasingly robust evidence. Previous research syntheses identified tutoring as among the most effective strategies for reading (Slavin et al. 2011), and updated syntheses have found similar results (Inns, Lake, Pelligrini, & Slavin, 2018). The current findings contribute to a smaller but growing evidence-base for the promise of tutoring for math outcomes (Pelligrini et al., 2018). Although sufficient research has been conducted on volunteer-based literacy tutoring programs to conduct meta-analytic research with positive findings for effects on reading and writing outcomes (Ritter et al., 2009), additional individual studies are necessary in math. Relatedly, research for community-supported math interventions needs to be conducted with different age and grade levels, including a focus on understanding the role and potential for community-based interventions to support early math outcomes (Mazzocco & Myers, 2003). In addition to research on the direct effects of community-based interventions on math outcomes, research is necessary to better understand the variables and conditions of effective inter-organizational partnerships for math interventions (Aarons et al., 2011). Such research holds potential to support efficacious implementation as well as identify ways to sustain implementation in a cost-effective manner long-term (e.g., Hollands et al., 2016).

Limitations

The contributions of this study should be considered in context of the limitations when interpreting results from the present study. First, the present evaluation was of a specific math intervention program. It is unclear if similar programs or subcomponents of Math Corps would produce comparable results. Second, it is common for math researchers to include a direct assessment of the set of skills targeted during intervention (e.g., Jitendra et al., 2015; Poncy et al., 2007; Witzel et al., 2003), yet no such measure was included in the current study. Math Corps provides support for a wide range of skills that vary across grades. Although we might have created an assessment directly aligned with the intervention program for each grade, the procedures for test creation were considered to be too burdensome and not of long-term relevance to either educator or policymaker stakeholders interested on grade-level performance. Further, the use of STAR Math and the state test is likely a more rigorous approach to estimating intervention effects than an assessment designed to assess the exact content included for intervention.

It is also important to note that the randomization procedures resulted in baseline differences in STAR Math scores between students assigned to Math Corps and students assigned to the control group. In addition to our final approach to analysis, which included trimming cases outside of two standard deviations of the mean, we conducted a series of sensitivity tests that included making no modifications to the analytic model, propensity score weighting, and fitting a model with no adjustment for pre-test scores (i.e., placing students assigned to Math Corps at a disadvantage). These sensitivity analyses showed no substantive changes to statistical significance and the corresponding effects and thus we opted to retain the trimmed model which eliminated baseline differences. Notably, the impact of addressing the students with the lowest scores is an important practical benefit to the schools that were served by Math Corps and suggest that the results favoring Math Corps may actually be more conservative given that the control group had higher initial scores. There were also a number of significant associations observed between demographic covariates included in the model and the outcomes of interest. Although those associations were not a focus of the current study, future research may continue to evaluate the relationship between demographic characteristics such as grade, gender, and race with intervention outcomes. Such work might provide a more nuanced view of intervention effects and offer promising avenues for improving school-based services.

The current study aimed to evaluate an existing community-supported math intervention program as it was implemented in practice. However, it is important to note that future research may advance the present study by examining a greater number of schools and students and conducting a priori power analyses (e.g., the power of the logistic regression models in the present study was not evaluated a priori). It may also be useful to evaluate or improve facets of the program. For example, the informal inventory adopted by tutors to inform progress within the intervention curriculum may benefit from more rigorous psychometric evaluation. Many students in the current study did not advance to grade-level standards and the inventory adopted by the program may have inhibited progress from one lesson to the next. Relatedly, the results observed in the present study should be interpreted within the context of the specific intervention evaluated and student population served. Different results may have been observed if a different Tier 2 intervention were used or if a broader population of students were included. However, during the year of the study, Math Corps tutors were placed at 105 different schools within the state of Minnesota. The demographics, as previously noted, and fall achievement data for participating students was generally commensurate with the profile of all Math Corps students across the state. The dosage received by participating students (approximately 22 weeks of support with an average of 69 min each week) was highly similar to all Math Corps students who received and average of 20 weeks of support (65 min each week). Finally, there were no statistically significant differences between the sample of participating students and the full Math Corps population when examining fall and spring STAR Math scores. The alignment between the study sample and broader Math Corps population provides some evidence for the generalizability of the results to the broader set of schools served by the program.

Conclusion

Improved math outcomes are needed for more students to experience math success and the accompanying the educational and life benefits. The current study suggests Math Corps can improve foundational math skills necessary for working with whole and rational numbers, but additional research is needed to extend those benefits to societally-endorsed outcomes like state proficiency tests. Nonetheless, this study adds to the general research base for both evidence-based math interventions and the delivery of such interventions via community-based resources.

References

Aarons, G. A., Hurlburt, M., & Horwitz, S. M. (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health and Mental Health Services Research*, *38*(1), 4-23.

Authors. (2019). Evaluation of a math intervention program implemented with community support. *Journal of Research on Educational Effectiveness*, *12*(3), 391-412.

Bailey, D. H., Siegler, R. S., & Geary, D. C. (2014). Early predictors of middle school fraction knowledge. *Developmental Science*, *17*, 775-785.

Bambara, L. M., Nonnemacher, S., & Kern, L. (2009). Sustaining School-Based Individualized Positive Behavior Support: Perceived Barriers and Enablers. *Journal of Positive Behavior Interventions*, *11*, 161–176.

Balu, R., Zhu, P., Doolittle, F., Schiller, E., Jenkins, J., & Gersten, R. (2015). Evaluation of response to intervention practices for elementary school reading. NCEE 2016-4000. *National Center for Education Evaluation and Regional Assistance*.

Berch, D. B., E., & Mazzocco, M. M. M. (2007). *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities*. Brookes Publishing Company.

Bosworth, K., Gingiss, P. M., Potthoff, S., & Roberts-Gray, C. (1999). A Bayesian model to predict the success of the implementation of health and education innovations in school-centered programs. *Evaluation and Program Planning*, *22*, 1–11.

Bronfenbrenner, U. (1987). *The Ecology of Human Development*. Barcelona: Polity Press.

Burns, M. K. (2005). Using incremental rehearsal to increase fluency of single-digit multiplication facts with children identified as learning disabled in mathematics computation. *Education & Treatment of Children (West Virginia University Press)*, *28*(3), 237–249.

Burns, M. K., Jimerson, S. R., VanDerHeyden, A. M., & Deno, S. (2016). Toward a Unified response-to-intervention model: multi-tiered systems of support. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of multi-tiered systems of support* (2nd ed.; pp. 719-732). New York, NY: Springer.

Burns, M. K., Ysseldyke, J., Nelson, P. M., & Kanive, R. (2015). Number of repetitions required to retain single-digit multiplication math facts for elementary students. *School Psychology Quarterly*, *30*(3), 398.

Burns, M. K. (2005). Using incremental rehearsal to increase fluency of single-digit multiplication facts with children identified as learning disabled in mathematics computation. *Education and Treatment of Children*, 237-249.

Claessens, A., & Engel, M. (2013). How important is where you start? Early mathematics knowledge and later school success. *Teachers College Record*, *115*(6), 1– 29.

Corporation for National and Community Service (2018). *Volunteering in U.S. hits record high; Worth $167 Billion.* Author. Retrieved December 31, 2018 from https://www.nationalservice.gov/newsroom/press-releases/2018/volunteering-us-hits-record-high-worth-167-billion.

DeFouw, E. R., Codding, R. S., Collier-Meek, M. A., & Gould, K. M. (2018). Examining dimensions of treatment intensity and treatment fidelity in mathematics intervention research for students at risk. *Remedial and Special Education*, online first.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., … Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*, 327-350.

Foegen, A. (2000). Technical adequacy of general outcome measures for middle school mathematics. *Diagnostique*, *25*, 175–203.

Foegen, A., & Deno, S. L. (2001). Identifying growth indicators for low-achieving students in middle school mathematics. *Journal of Special Education*, *35*, 4–16.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*, 93-98.

Fuchs, D., & Fuchs, L. S. (2017). Critique of the national evaluation of response to intervention: A case for simpler frameworks. *Exceptional Children*, *83*, 255-268.

Fuchs, L. S., Powell, S. R., Cirino, P. T., Schumacher, R. F., Marrin, S., Hamlett, C. L., ... Changas, P. C. (2014). Does calculation or word-problem instruction provide a stronger route to pre-algebraic knowledge? *Journal of Educational Psychology, 106,* 990–1006.

Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-intervention: A decade later. *Journal of learning disabilities*, *45*(3), 195-203.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive Mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78*(4), 1343–1359.

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. (2009). Assisting students struggling with mathematics: Response to Intervention (RtI) for elementary and middle schools. NCEE 2009-4060. *What Works Clearinghouse*.

Graham, J. Olchowski, A., & Gilreath, T. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, *8*(3), 206–213.

Hollands, F. M., Kieffer, M. J., Shand, R., Pan, Y., Cheng, H., & Levin, H. M. (2016). Cost-Effectiveness Analysis of Early Reading Programs: A Demonstration With Recommendations for Future Research. *Journal of Research on Educational Effectiveness*, *9*(1), 30-53.

Inns, A., Lake, C., Pellegrini, M., & Slavin, R. (2018). *Effective programs for struggling readers: A best-evidence synthesis.* Paper presented at the annual meeting of the Society for Research on Educational Effectiveness, Washington, DC.

Jacob, R., Armstrong, C., Bowden, A. B., & Pan, Y. (2016). Leveraging volunteers: An experimental evaluation of a tutoring program for struggling readers. *Journal of Research on Educational Effectiveness*, *9*(1), 67-92.

Jimerson, S. R., Burns, M.K. VanDerHeyden, A. M. (2016). From response to intervention to mulit-tiered systems of support: Advances in the science and practice of assessment and intervention. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of multi-tiered systems of support* (2nd ed.; pp. 1-6). New York, NY: Springer.

Jitendra, A. K., Lein, A. E., Soo-hyun Im, Alghamdi, A. A., Hefte, S. B., & Mouanoutoua, J. (2018). Mathematical Interventions for secondary students with learning disabilities and mathematics difficulties: A Meta-Analysis. *Exceptional Children*, *84*(2), 177–196.

Jitendra, A. K., Petersen-Brown, S., Lein, A. E., Zaslofsky, A. F., Kunkel, A. K., Jung, P.-G., & Egan, A. M. (2015). Teaching Mathematical word problem solving: The quality of evidence for strategy instruction priming the problem structure. *Journal of Learning Disabilities*, *48*(1), 51–72.

Jitendra, A. K., Sczesniak, E., & Deatline-Buchman, A. (2005). An exploratory validation of curriculum-based mathematical word problem-solving tasks as indicators of mathematics proficiency for third Graders. *School Psychology Review*, *34*(3), 358–371.

Jordan, N. C., Hansen, N., Fuchs, L. S., Siegler, R. S.,Gersten, R., & Micklos, D. (2013). Developmental predictors of fraction concepts and procedures. *Journal of Experimental Child Psychology*, *116*, 45–58. doi:10.1016/j.jecp.2013.02.001

Jordan, N.C., Resnick, I., Roriguez, J. Hansen, N., & Dyson, N. (2017). Delaware longitudinal study of fraction learning: Implications for helping children with mathematics difficulties. *Journal of Learning Disabilities, 50*, 621-630.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2004). *Applied linear statistical models* (5th ed.). Boston: McGraw-Hill/Irwin.

Kratochwill, T. R. (2007). Preparing psychologists for evidence-based school practice: Lessons learned and challenges ahead. *American Psychologist*, *62*, 829-843.

Long, A. C. J., Sanetti, L. M. H., Collier-Meek, M. A., Gallucci, J. J., Altschaefl, M., & Kratochwill, T. R. (2016). An exploratory investigation of teachers' intervention planning and perceived implementation barriers. *Journal of School Psychology, 55,* 1-26.

Markovitz, C.; Hernandez, M.; Hedberg, E.; Silberglitt, B. (2014). *Impact evaluation of the Minnesota reading corps k-3 program*. NORC at the University of Chicago: Chicago, IL.

Mazzocco, M. M. M. & Myers, G. F. (2003). Complexities in identifying and defining mathematics learning disability in the primary school-age years. *Annals of Dyslexia, 53,* 218-253.

Minnesota Department of Education (2017). *Procedures manual for the Minnesota assessments 2017-2018.* Retrieved from http://minnesota.pearsonaccessnext.com/resources/resources-training/manuals/2017-18_Procedures_Manual.pdf

Montague, M. (1997). Cognitive strategy instruction in mathematics for students with learning disabilities. *Journal of learning disabilities, 30*(2), 164-177.

Morgan, P. L., Farkas, G., & Wu, Q. (2011). Kindergarten children's growth trajectories in reading and mathematics: Who falls increasingly behind? *Journal of Learning Disabilities, 44*(5), 472–488.

National Center for Education Statistics. (2017). *National assessment of educational progress: Mathematics.* Retrieved from *https://www.nationsreportcard.gov/math_2017.*

National Mathematics Advisory Panel (2008). *Foundations for Success: The Final Report of the National Math Advisory Panel.* U.S. Department of Education: Washington, DC.

Nelson, P. M., Parker, D. C., & Zaslofsky, A. (2016). The relative value of growth in math fact skills across late elementary and middle school. *Assessment for Effective Intervention, 41,* 184-192.

Nelson, P. M., Parker, D. C., & Van Norman, E. R. (2018). Subskill mastery among elementary and middle school students at risk in mathematics. *Psychology in the Schools*, *55*(6), 722–736.

Noell, G. H., & Gansle, K. A. (2016). Assuring the response to intervention process has substance: Assessing and supporting intervention implementation. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of multi-tiered systems of support* (2nd ed.; pp. 407-420). New York, NY: Springer.

Parsons, S., & Bynner, J. (1997). Numeracy and employment. *Education and Training, 39,* 43–51.

Pellegrini, M., Inns, A., & Slavin, R. (2018). *Effective programs in elementary mathematics: A best-evidence synthesis.* Paper presented at the annual meeting of the Society for Research on Educational Effectiveness, Washington, DC.

Poncy, B. C., Skinner, C. H., & Jaspers, K. E. (2007). Evaluating and comparing interventions designed to enhance math fact accuracy and fluency: cover, copy, and compare versus taped problems. *Journal of Behavioral Education*, *16*(1), 27–37.

Renaissance Learning (2013). *STAR Math: Technical manual.* Wisconsin Rapids, WI: Author.

Ritter, G. W., Barnett, J. H., Denny, G. S., & Albin, G. R. (2009). The effectiveness of volunteer tutoring programs for elementary and middle school students: A Meta-Analysis. *Review of Educational Research*, *79*(1), 3–38.

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys* (Vol. 81). John Wiley & Sons.

Shalev, R. S., Manor, O., & Gross-Tsur, V. (2005). Developmental dyscalculia: A prospective six-year follow-up. *Developmental Medicine and Child Neurology, 47,* 121–125.

Shapiro, E. S., & Gebhardt, S. N. (2012). Comparing computer-adaptive and curriculum-based measurement methods of assessment. *School Psychology Review, 41,* 295–305.

Skinner, C. H., Turco, T. L., Beatty, K. L., & Rasavage, C. (1989). Cover, copy, and compare: A method for increasing multiplication performance. *School Psychology Review*, *18*, 412-420.

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.

Spectrum K12, American Association of School Administrators, Council of Administrators of Special Education, National Association of State Directors of Special Education, & State Title 1 Directors (2010). *Response to intervention (RTI) adoption survey.*

Stickney, E. M., Sharp, L. B., & Kenyon, A. S. (2012). Technology-enhanced assessment of math fact automaticity: Patterns of performance for low- and typically achieving students. *Assessment for Effective Intervention, 37,* 84–94.

Thurber, R. S., Shinn, M. R., & Smolkowski, K. (2002). What is measured in mathematics tests? Construct validity of curriculum-based mathematics measures. *School Psychology Review*, *31*, 498–513.

Travers, J.C., Cook, B.G., Therrien, W.J., & Coyne, M.D. (2016). Replication research and special education. *Remedial & Special Education*, *37,* 195-204.

Van Houten, R., & Thomas, C. (1976). The effects of explicit timing on math performance. *Journal of Applied Behavior Analysis*, *9*, 227—230.

Vujnovic, R. K., Fabiano, G. A., Morris, K. L., Norman, K., Hallmark, C., & Hartley, C. (2014). Examining school psychologists' and teachers' application of approaches within a response to intervention framework. *Exceptionality*, *22*, 129-140.

Vukovic, R. K., Fuchs, L. S., Geary, D. C., Jordan, N. C., Gersten, R., & Siegler, R. S. (2014). Sources of individual differences in children's understanding of fractions. *Child Development*, *85*, 1461–1476.

Wandersman, A., Duffy, J., Flaspohler, P., Noonan, R., Lubell, K., Stillman, L., ... & Saul, J. (2008). Bridging the gap between prevention research and practice: the interactive systems framework for dissemination and implementation. *American Journal of Community Psychology*, *41*, 171-181.

What Works Clearinghouse (2016). *Find what works.* Retrieved December 12, 2018 from https://ies.ed.gov/ncee/wwc/FWW/Results?filters=,Math.

Witzel, B. S., Mercer, C. D., & Miller, M. D. (2003). Teaching algebra to students with learning difficulties: An investigation of an explicit instruction model. *Learning Disabilities Research & Practice*, *18*(2), 121-131.

Table 1

Student Demographics across Conditions

Demographic Category	Math Corps (n = 416)	Control (n = 259)
Grade		
Fourth	17.5%	17.0%
Fifth	21.9%	20.5%
Sixth	29.6%	30.1%
Seventh	19.3%	19.3%
Eighth	13.1%	13.1%
Gender		
Male	48.8%	44.0%
Ethnicity		
White	53.4%	52.9%
Black	24.5%	23.2%
Latinx	10.6%	12.0%
Asian-American	8.9%	6.9%
Native North American	1.2%	0.5%
Other	0.7%	2.3%

Table 2.

Descriptive Outcomes by Assessment Period and Condition

		Fact Fluency			STAR Math			State Test		
		M	*SD*	Met Benchmark	*M*	*SD*	Met Benchmark	*M*	*SD*	Met Benchmark
Pre	Control	16.38	7.91	-	669.26	81.51	-	-	-	-
	Math Corps	16.99	0.45	-	659.47	89.71	-	-	-	-
Post	Control	19.46	8.58	11%	716.87	80.43	15%	41.47	8.33	15%
	Math Corps	23.29	8.31	23%	728.09	90.58	27%	42.19	8.73	19%

Table 3

Fixed Effects from Multi-Level Regression Models across Outcomes

	Fact Fluency		STAR Math		MCA	
Fixed Effects	B	SE	B	SE	B	SE
Intercept	19.05**	1.14	612.48**	8.56	41.97**	1.07
Prior Achievement	0.72**	0.04	0.54**	0.06	0.05**	0.01
Grade Level						
Fifth	0.69	0.97	65.64**	8.49	-1.52	1.05
Sixth	1.98	1.07	131.32**	8.54	-0.12	1.06
Seventh	-0.51	1.35	139.98**	10.34	1.15	1.27
Eighth	1.27	1.38	144.93**	10.27	1.97	1.29
Gender						
Boys	-0.47	0.59	12.03**	5.19	0.32	0.63
Race						
Black	-0.61	0.84	-12.82	7.02	-3.10**	0.86
Latino	-0.32	0.97	-5.14	8.58	-1.21	1.05
Other	1.95	1.02	9.84	9.15	-1.09	1.12
Group Assignment						
Math Corps	3.48**	0.65	16.63**	5.52	1.21	0.67

** $p < .01$

Note: Sample sizes for each model equal to those reported in Table 2. Prior Achievement = mean centered fall STAR Math score or mean centered fall fact fluency score. Data clustered at the tutor level. Due to low counts, Asian, Pacific Islander, Native American, and Multi-race groups combined into Other.

Table 4

Results of Generalized Linear Mixed Models Predicting Spring Benchmark Attainment on STAR Math and the MCA

	STAR Math		MCA	
Variable	Log-Odds	SE	Log-Odds	SE
Intercept	-1.89***	0.33	-1.32	0.33
Prior Achievement	0.02***	0.00	0.01**	0.00
Grade Level				
Fifth	-0.72*	0.34	-0.97*	0.38
Sixth	0.12	0.32	-0.36	0.34
Seventh	-1.30**	0.45	-0.51	0.43
Eighth	-0.90*	0.43	0.31	0.39
Gender				
Male	0.40	0.22	-0.03	0.22
Race				
Black	-0.40	0.30	-0.61	0.32
Latino	-0.36	0.39	-0.77	0.42
Other	0.25	0.35	-0.24	0.39
Group Assignment				
Math Corps	0.98***	0.25	0.38	0.24

* $p < .05$, ** $p < .01$, *** $p < .001$

Note: MCA = Minnesota Comprehensive Assessment in Math