ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 20-310

Engaging Girls in Math: The Unequal Effects of Text Messaging to Help Parents Support Early Math Development

Christopher Doss
RAND Corporation

Hans Fricke
Amazon.com, Inc.

Susanna Loeb
Brown University

Justin B. Doromal
Brown University

This study assesses the effects of two text messaging programs for parents that aim to support the development of math skills in prekindergarten students. One program focuses purely on math, while the other takes an identical approach but focuses on a combination of math, literacy, and social-emotional skills. We find no evidence that the math-only program benefits children's math development. However, the combination program shows greater promise, particularly for girls. Quantile regressions indicate that the effects are concentrated in the lower half of the outcome distribution. We discuss and provide evidence for various hypotheses that could explain these differences.

VERSION: April 2022

Suggested citation: Doss, Christopher, Hans Fricke, Susanna Loeb, and Justin B. Doromal. (2022). Engaging Girls in Math: The Unequal Effects of Text Messaging to Help Parents Support Early Math Development. (EdWorkingPaper: 20-310). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/39zc-j672

Engaging Girls in Math:
The Unequal Effects of Text Messaging to Help Parents Support Early Math Development

Christopher Doss[a]
RAND Corporation

Hans Fricke[b]
Amazon.com, Inc.

Susanna Loeb[c]
Brown University

Justin B. Doromal[d]
Brown University

Abstract: This study assesses the effects of two text messaging programs for parents that aim to support the development of math skills in prekindergarten students. One program focuses purely on math, while the other takes an identical approach but focuses on a combination of math, literacy, and social-emotional skills. We find no evidence that the math-only program benefits children's math development. However, the combination program shows greater promise, particularly for girls. Quantile regressions indicate that the effects are concentrated in the lower half of the outcome distribution. We discuss and provide evidence for various hypotheses that could explain these differences.

Highlights:

- We test the effects of two text messaging curricula that leverage behavioral economics principles to help parents support the math development of prekindergarteners in the home.
- We find that a program that cycles through literacy, mathematics, and social-emotional skills increases math achievement for girls, while a program focusing solely on mathematics has no effects.
- Benefits for girls are concentrated on those with weaker performance on mathematics assessments.
- We posit potential mechanisms based on the literature.

Keywords: Educational Economics, Human Capital, Behavioral Economics, Early Childhood Education, Randomized Controlled Trials

JEL: I21, I24

Declarations of interest: None

[a] Corresponding Author; 1200 South Hayes Street, Arlington, VA 22202, cdoss@rand.org
[b] Terry Avenue North, Seattle, WA, 98109, hans.fricke@gmail.com. Hans Fricke worked on this project before joining Amazon.
[c] 164 Angell Street, Providence RI, loeb@brown.edu
[d] 164 Angell Street, Providence RI, justin_doromal@brown.edu

Introduction

Math education is almost always on the forefront of education policy discussions. Interest in math from politicians, policymakers, researchers, and practitioners stems, in part, from the understanding that the economy depends on jobs in science, technology, engineering, and mathematics (STEM), which all require math skills. The United States' relatively poor performance on international tests has fueled fears that the nation will lose its competitive advantage (Committee on STEM Education, 2018; Clements, 2004). This emphasis on math has more recently reached the early childhood education sector where more and more children are exposed to early math concepts in prekindergarten (Clements, 2004).

Research provides some evidence that starting math instruction early is beneficial to children's development of math skills. Math is cumulative; more advanced concepts such as addition and subtraction depend on knowledge of earlier concepts such as number recognition and counting. Math achievement in prekindergarten and kindergarten predicts future math achievement as late as eighth grade (Claessens and Engel, 2013; Jordan et al., 2009). More generally, investments in quality early childhood learning experiences provide some of the highest returns on investment in education (Heckman, 2006).

While math is increasingly taught in formal early childhood education settings, young children's exposure to early math is less prevalent at home. Parents tend to endorse the notion of supporting early math concepts in the home, but they focus more on literacy development than math development, with some evidence that this preference for literacy is greater among low-income, black, and Hispanic families (Sonnenschein Metzger, and Thompson, 2016; Sonnenschein et al., 2012; Cannon and Ginsburg, 2008). Furthermore, studies have shown that low-income families are less likely than higher-income families to foster the academic development of their children at home more broadly (Sonnenschein Metzger, and Thompson, 2016; Bradley et al., 2001). Home learning environments are consequential for children. Differences in home experiences early on in life affect early child development and persist through school and beyond (Duncan, Ziol-Guest, and Kalil, 2010; Pungello et al., 2010).

Text messaging interventions for parents have proven to promote behaviors that improve educational outcomes in a variety of settings. In particular, text messaging curricula including information, activities, and encouragement have helped parents support the literacy development of their prekindergarteners in the home. These programs were most effective for students who started the year with lower literacy skills (Doss et al., 2019; York, Loeb, and Doss, 2019; Cortes et al., 2021; and Cortes et al., 2019).

To date, no study that we know of has assessed the effects of text messaging programs – or similar low-touch programs – for parents on children's development of math skills. The acquisition of math skills may differ from the acquisition of literacy skills for a number of reasons. In particular, parents are primed to focus on early literacy. Programs such as "Talk Read Sing" and "Reach Out and Read" encourage parents to build their children's early literacy skills (Nobles, W.P., 2018; Mendelsohn et al., 2001). Very few similar signals instruct parents to focus on math. Moreover, parents may be more comfortable thinking of themselves as teachers of literacy than teachers of math. While early math is likely no more difficult for parents – only requiring very basic understanding of counting, numbers and comparisons such as more and less – parents may be worried about math because their own experience with math in schools may not have been positive, and they may experience math anxiety (Ashcraft and Moore, 2009; Luttenberger, Wimmer, and Paechter, 2018). Finally, math may elicit gender-specific behaviors in different ways than literacy. Math is traditionally considered a male subject area (Gunderson et al., 2012; Cvencek and Greenwalk, 2011). As such parents' inclination to build math skills may differ depending on their child's gender.

In this study, we investigate whether 32-week, text messaging-based programs can help parents improve the early math outcomes of their prekindergarteners. We field an experiment to compare a program that focuses solely on math ("pure math program") with one that combines math, literacy and social-emotional learning (SEL) ("combination program") in three California school districts. The former may be more effective at improving math because it spends more time on math (Berkowitz et al., 2015). The latter may be more effective because it combines literacy, for which parents are primed to work and generally feel comfortable, and because literacy and social-emotional skills may enable children to more readily acquire math skills (Purpura et al. 2011; Graziano, 2007). Both programs send three text messages per week to promote behavior change of parents and to foster positive parent-child interactions similar to those developed by York et al. (2019).

We find that the pure math program had no detectable effects on assessments of early math skills for either girls or boys. The combination program had meaningful positive effects on girls of 0.16 standard deviations, and quantile regression analysis indicates that girls between approximately the 15th and 55th percentile of the outcome distribution have large and significant benefits of approximately 0.30 to 0.40 standard deviations. The magnitude of the effect implies that while girls start off the year behind boys in math, they catch up to and even surpass boys when their parents have access to the program that combines literacy, math, and SEL. We hypothesize three potential explanations of these differences. First, parents of boys may be overconfident in boys' ability to do math and see less need to practice math with boys. Second, parents and teachers may respond to greater (perceived) girls' interest in and mastery math related activities. Third, while parents may not engage differently based on their child's gender, girls may be more receptive to the math activities because they can regulate emotions and impulses better than boys (Matthews, Ponitz, and Morrison, 2009; Ponitz et al., 2008; Else-Quest et al., 2006).

Background

A. Early Childhood Math

A substantial body of research shows that early childhood math skills predict later mathematics outcomes. For example, knowledge at age four of specific concepts such as group size, counting, and pattern recognition predicts fifth grade mathematics achievement. This relationship is mediated by first grade math knowledge, in support of the hypothesis that early math knowledge creates a beneficial learning trajectory (Rittle-Johnson et al., 2017). Math achievement in kindergarten predicts both the *rate* of math growth through third grade as well as the math achievement of children in third grade (Jordan et al., 2009). More generally, an analysis of the Early Childhood Longitudinal Study, Kindergarten Cohort 1998 (ECLS-K 98) shows that kindergarten math knowledge predicts math, literacy, and science achievement, as well as grade retention, and that math knowledge is more predictive of later academic achievement than kindergarten literacy knowledge (Claessens and Engel, 2013).

While math knowledge predicts later school success, parents often ignore or deemphasize math learning for young children in the home. Though parents conceptually endorse the notion of supporting early math concepts at home, they tend to focus more on literacy development than math development (Sonnenschein Metzger, and Thompson, 2016; Sonnenschein et al., 2012; Cannon and Ginsburg, 2008), in part, because parents tend to see math instruction as the responsibility of schools (Gunderson and Levine, 2011) and, perhaps, in part due to their own math anxiety (Ashcraft and Moore, 2009; Luttenberger, Wimmer, and Paechter, 2018). This tendency to focus less on math development may be a detriment to the child, as the number of home parent math activities and even the amount of home math talk when the child is four years old and younger are correlated with early math performance (Manolitsis, Georgiou, and Tziraki, 2013; Gunderson and Levine, 2011; Pruden, Levine, and Huttenlocher, 2011; Levine et al., 2010).

This lack of emphasis on math may be particularly detrimental for students from low-income families. These students exhibit weaker math skills at the time children enter preschool at the age of 4 which, at least in part, likely stem from differential inputs in the home (Claessens and Engel, 2013; Jordan et al., 2009; Clements, 2004). Indeed, studies illustrate that low-income families are less likely to foster academic development of their children at home. While parents of all economic levels engage in math talk in the home, lower-income families are more likely to focus on more basic concepts such as counting, while higher income families include more advanced concepts such as comparison of set sizes, cardinality of numbers, addition, and subtraction (Levine et al., 2010). Further, research provides some evidence that the preference to foster literacy development over math development in the home is greater for low-income families (Sonnenschein Metzger, and Thompson, 2016), a pattern that potentially contributes to the achievement gap in mathematics emerging by the age of four.

B. Text Message-Based Behavioral Interventions

Behavioral economics provides at least four insights as to why parents may underinvest in the academic development, including the math development, of their children in the home. First, people tend to underinvest in areas if they have incomplete information on the benefits of those investments (Mullainathan and Thaler, 2000). Research on how incomplete information affects educational outcomes is mixed, with some studies finding that more information affects behavior and other studies finding no effects (Fricke, Grogger, and Steinmayr 2018; Rogers and Feller, 2016; Kraft and Rogers, 2015; Valant and Loeb, 2014; Hastings and Weinstein, 2008; Grodsky and Jones, 2007; Avery and Kane, 2004). With respect to home practices, parents may be unaware of the longer-term benefits of developing academic skills in the home or which skills to develop. Given the relatively recent emphasis on early math, and the more traditional emphasis on early literacy, lack of information may be especially salient in the math context.

Second, adults also tend to underinvest in cognitively complex tasks; the underinvestment is potentially larger for people who deal with cognitively demanding challenges of poverty, health issues or other life stressors (Mani et al, 2013; Shah, Mullainathan, and Shafir, 2012; Iyengar and Lepper, 2000). The perceived cognitive complexity of building math skills may be even greater compared to other subjects such as literacy because many parents are unaware of *how* to build the math skills of their child in the home (Gunderson and Levine, 2011). Further, the cognitive barrier to math education may be uniquely higher due to math anxiety. Researchers have estimated that up to 20 percent of the population has math anxiety that is strong enough to lead them to withdraw from math-oriented activities (Eden, Heine, and Jacobs, 2013; Hembree, 1990). Parent math anxiety can hamper the math development of their children (Maloney et al., 2015).

Third, parents must also build academic skills in the home while juggling a multitude of other professional and personal responsibilities. Parenting is not a one-time activity. They have to remember to create positive learning opportunities for their children on an on-going basis. While parents may have the intention of building academic skills in their children at home, they may forget to do so with so many competing demands on their attention, and because attention is limited (c.f., Karlan et al., 2016). Fourth, this effect may be exacerbated by the fact that the benefits of their efforts may not be seen for many months or years. With time-inconsistent preferences, parents may discount the future benefits of these activities and underinvest in the present (DellaVigna, 2009).

Text messaging has been proven to be a useful medium by which interventions can be administered to address these behavioral barriers and change behavior. Texting has been successfully used at all levels of education from helping parents enroll their children in prekindergarten (Weixler et al., 2020), to reducing chronic absenteeism and increasing parental engagement (Smythe-Leistico and Page, 2018) in prekindergarten, to positively affecting school and class attendance (Bergman and Chan, 2017; Groot, Sander, Rogers, and Bloomenthal, 2017; Robinson, Lee, Dearing, and Rogers 2017; Rogers and Feller 2018), to promoting assignment completion (Bergman, 2015), to helping students pass summer school (Kraft and Rogers, 2015), to helping students complete FAFSA (Page, Castleman, and Meyer, 2016) and enroll into college (Castleman and Page, 2015, 2016).

Most pertinent to this study, text messaging curricula have been used to help parents support the literacy development of their prekindergarteners in the home (Doss et al., 2019; York, Loeb, and Doss, 2019; Cortes et al., 2021; and Cortes et al., 2019). These curricula send parents three text messages a week, each to designed to address one or more behavioral barriers. A “FACT” text sent on Mondays explains to parents the skill of the week and why it is important, addressing potential information asymmetries. A “TIP” text on Wednesdays breaks down the cognitive complexity of building literacy by providing a small, easy-to-implement activity that leverages everyday objects and routines to build the skill of the week. Finally, a “GROWTH” text on Fridays serves as a reminder of the benefits of the program, to prevent parents from discounting the future benefits of the activities. All texts also serve as reminders to overcome limited attention.

Encouragement such as "You are preparing your child 4 K" are intended to overcome the problem of delayed gratification in combination with time inconsistent preferences.[1]

As these text messaging programs proliferate, it is becoming more evident that the structure of the programs can have dramatic implications for their effectiveness. York, Loeb, and Doss (2019) find that their literacy curriculum, which provides all parents with the same text messages, is most effective for children in the bottom half of the baseline skills distribution, perhaps because the difficulty of the activities is best aligned to children at that level. Doss et al. (2019) show that extra learning gains can be extracted by aligning the difficulty of the task to the ability of the child. A text messaging program that provides harder tips to parents of kindergarteners who score higher on formative assessments was more effective than a program that gave the same tips to all parents. Evidence suggests that those extra gains were concentrated on children at the top of the baseline skills distribution – the segment of children least served by the general texting program.

Additionally, there is suggestive evidence that texting programs that cycle through literacy, math, and social-emotional skills are more effective at raising literacy skills than an equally long texting program that focuses solely on literacy development (York, Loeb, and Doss, 2019). The authors hypothesize that cycling through domains may keep the parents more engaged in the program, and the diversity of academic domains may keep parents engaged if they struggle on one domain but find success in another. Moreover, these domains are not mutually exclusive at this early age and may be complementary (Butterworth 2005; Graziano et al. 2007; Sarama et al.2012; Morris et al. 2013).

Finally, subgroups of populations are differentially sensitive to alterations to the three-message format (FACT, TIP, GROWTH) and the days on which those messages are sent. Cortes

[1] Such stimuli represent self-affirmation, which has been found to be effective in the behavioral science literature (Sweeney and Moyer, 2015; Cohen and Sherman, 2014; Hall, Zhao, and Shafir, 2014).

et al. (2021) show that frequency of text messages matters for program effectiveness and depends on students' baseline literacy skills. While one "TIP" text per week seems to be most effective for students in the middle of the baseline literacy distribution, students in the lowest quarter benefit most from three texts ("FACT", "TIP", and "GROWTH"). Five texts with additional activities do not improve literacy development, but appear to dampen parents' program experience and lead parents to opt out of the program at higher rates. Cortes et al. (2019) provide evidence that parenting support works best when parents have time, attention, and need. They find that sending text messages on weekends is more beneficial to children's development than sending texts on weekdays. The benefits of the weekend texts were particularly pronounced for children who started prekindergarten in the lower half of the baseline skill distribution on easier subcomponents, while the weekday texts were somewhat more beneficial for the initially higher achieving children. Fricke, Kalogrides, and Loeb (2018) provide evidence that text messages with more complex language and programs with only activities, compared to programs that scaffold activities with context and encouragement, increase the likelihood that parents opt out.

These findings illustrate that, in addition to studies that test new ways to leverage text messaging and behavioral economics to improve a broader range of outcomes, attention directed at understanding how the design of interventions affects impact could be beneficial. This study contributes to both those aims. We extend the work of York, Loeb, and Doss (2019) to understand whether parent-facing text messaging curricula can improve the early childhood math outcomes of prekindergarteners while directly testing whether a program focused solely on math or a program that cycles through literacy, math, and social-emotional skills is more effective at raising math achievement.

C. The Intervention

The intervention tests the effects of two different text messaging programs designed to support the development of preschool-aged students: the *pure math* program and the *combination* program. Both programs were created by our research team and are based on the literacy-focused text messaging program first fielded in San Francisco Unified School District in the 2013-14 school year (York et al., 2019). The programs lasted eight months (32 weeks) from mid-October to mid-June. Similar to the original program, the pure math and the combination program follow the “FACT”, “TIP”, and “GROWTH” approach to provide information and alleviate behavioral barriers to parenting.

The contents of the text messages differ between the programs. The pure math program focuses exclusively on building math related skills. These skills include counting, number recognition, shapes, sorting, patterns, addition, subtraction, and comparisons of size. The combination program covers literacy and SEL in addition to math topics. The topic rotates each week from literacy to math to SEL. Literacy texts cover upper- and lower-case letter recognition, letter-sound awareness, beginning-sound awareness, rhyme awareness, name writing, concepts of print, story comprehension, vocabulary development, listening to and singing songs, self-narration, parent-child conversations, and establishing high-quality parent-child book reading routines. SEL texts concentrate on identifying emotions, identifying their causes and consequences, building emotion regulation, perseverance, sharing, and turn-taking. In each program, we sent texts with skills and activities relevant for three-year-old and four-year-old students based on the California Preschool Learning Foundation age-specific standards. The texts also align with the Common Core standards and are based on research in academic development (e.g., Lonigan and Shanahan 2009; California Department of Education, 2008) and academic-related parenting practices (e.g., Reese, Sparks, and Leyva, 2010).[2]

We designed the text messages to support general positive parenting practices beyond focusing on certain skills. In addition, the text messages build on daily routines such as meals or bath time to make it as easy as possible for parents to implement the suggested activities. Finally, the programs employ a spiral curriculum, and as children grow older and develop, the activities become increasingly more advanced and repeat topics for reinforcement. The following texts are examples for each topic area:

Literacy:

> *FACT: Letters are the building blocks of written language. Children need to know the letters to learn how to read and write.*
>
> *TIP: Point out the first letter in your child's name in magazines, on signs & at the store. Have your child try. Make it a game. Who can find the most?*
>
> *GROWTH: Keep pointing out letters. You're preparing your child for K! Point out each of the letters in your child's name. Ask: What sound does it make?*

Math:

> *"FACT: Shapes are all around us. You can help build your child's math skills by pointing out shapes and asking questions about them."*
>
> *"TIP: Look for shapes on the go. Point & say: That house's windows are rectangles. Ask: What shape are the wheels on that car?"*
>
> *"GROWTH: Keep pointing out shapes. You're preparing for K! Make it a game. Who can find a circle, square, rectangle, and triangle? (like a slice of pizza)"*

SEL:

> *"FACT: Letting your child know that you see her/him trying makes her/him try harder. Talk about her/his effort, not on the end result."*
>
> *"TIP: Ask your child to help with a difficult task and talk about his/her effort: You worked hard to make your bed! The corners are tough!"*

[2] See http://www.cde.ca.gov/sp/cd/re/documents/preschoollf.pdf for the California standards and http://www.corestandards.org/wp-content/uploads/ for Common Core standards.

"GROWTH: Keep acknowledging your child's effort to prepare him/her for K! Even if your child didn't finish, say: You worked hard. That's important."

We fielded the intervention among parents of preschoolers in three school districts in California – San Francisco Unified School District, Oakland Unified School District, and Fresno Unified School District – during the school year 2017-18. We asked parents to participate as part of the prekindergarten enrollment process. Parents could choose between texts in English, Spanish, or Cantonese. We assigned parents into three equal-sized groups in a blocked randomization based on prekindergarten center and preferred texting language. The two treatment groups received either the pure math or the combination program, and the control group received a placebo program – one text message with school related information such as events or vaccinations every two weeks. The placebo program did not provide any information about child development or parenting practices.

Prior to this study, during the 2015-2016 school year, we tested both programs separately in two pilots in California, the pure math program in Jumpstart of Northern California and San Mateo Head Start Centers and the combination program in San Francisco Unified School District.[3] While neither of the programs had a significant effect for the full sample on students' math scores at the end of the school years, we found that the combination program increased girls' math scores significantly and meaningfully (approximately a third of a standard deviation). See Table 1 for results. These pilots have two obvious caveats. First, the lack of any significant effect for the math-only program could be due to insufficient power to detect smaller effects. Second, we fielded the programs in distinct districts and therefore students whose parents received the pure math programs and the combination program likely differ. As such, the differentially positive effect of

[3] For details of the experiments see Appendix A.

the combination program may be due to a more conducive sample. Overall, the results from these previous studies serve as a motivation to compare the math-only program directly to the combination program and to examine gender differences in effects specifically.

Table 1: Effect of 2015-2016 Math and Combination Pilot Texting Programs

	Standardized Math Score			
	Math Only	N	Combination	N
Partnering Site	Jump Start and San Mateo Head Start		SFUSD	
All Students	-0.011 (0.096)	523	0.104 (0.079)	418
Girls	-0.001 (0.115)	254	0.332* (0.130)	206
Boys	-0.076 (0.145)	269	-0.097 (0.120)	212

Notes: Math score is the average of the standardized subscores. The average is standardized to have mean zero and standard deviation one. All regression models include randomization block fixed effects and a full set of covariates. Standard errors are clustered on the randomization block. * indicates p<0.05.

Data

We draw on three main data sources for this study. First, to measure math development, we assessed participating students with the math section of the Brigance Inventory of Early Development III Standardized assessment in late April and May of 2018. The assessment is a validated one-on-one math assessment covering tasks related to skills highlighted in our text messaging programs.[4] We trained 43 assessors who visited the preschools and assessed the children individually for approximately 20 minutes. In particular, students were asked to (1) identify single digit amounts, (2) count as high as they can, (3) compare different amounts, (4) sort objects by color, size, and shape, (5) match quantities with numerals, (6) read numerals, (7) solve word problems such as “Are there enough balls (in the picture) so that every dog can have a ball?”, and (8) identify missing numerals in sequences. For each set of tasks, we counted the number of correct answers with the exception of rote counting, for which we used the highest number students

[4] For more information about the assessment, see https://www.curriculumassociates.com/products/brigance/early-childhood.

counted to correctly. We only included students into our sample who attempted at least one task on the math assessment.

Second, we received administrative student records from the participating districts. These data include student gender, student race and ethnicity, age, and scores of the fall Desired Results Developmental Profile (DRDP). The DRDP is a validated teacher observational assessment tool to assess five domains of child development: (1) approaches to learning, (2) social and emotional development, (3) language and literacy development, (4) math and science, and (5) physical development and health. The classroom teacher completes the DRDP for each of the children in the classroom over a two-week period. For each DRDP domain, the number of levels vary depending on the competencies that are appropriate for that domain's developmental continuum and are organized under the following four categories: (1) Responding, (2) Exploring, (3) Building, (4) Integrating.

Third, we collected parent-specific information on the enrollment forms. These data include demographic information such as parental education, income, and hours worked, as well as questions about parenting practices. Specifically, we asked parents to indicate their agreement with a statement that they had knowledge of how to support their child's math and literacy development and behavior on a five-point scale ("strongly disagree" to "strongly agree"). We also asked parents how often in a week, on a four-point scale ("not at all" to "more than four times"), they engage in specific parent-child activities related to math, literacy, and SEL; sample items include counting to 20 or higher, practicing rhyming, and talking about feelings. This enrollment form data has a higher prevalence of missing data than the other data sources because of two reasons. First, some parents chose not to answer these questions when filling out the form. Second, some parents of four-year old students had participated in a pilot text messaging program aimed to support SEL of three-year-old students in the previous year in SFUSD. These parents signed a consent form that applied to both years and therefore did not answer the questions on the enrollment form in this year.

Our sample consists of a diverse student population. Table 2 shows descriptive statistics for student and parent characteristics. Thirty-nine percent of students are Hispanic, 34 percent are Asian, 14 percent are black, and seven percent are white. Fifty percent of students are female, and students are on average 4.1 years old at enrollment into prekindergarten. Of parents who filled out the enrollment form and provided education information, 19 percent do not have a high school degree, 31 percent have a high school degree, 25 percent spent some time in college without obtaining a degree, eight percent have an Associate's degree, 11 percent have a Bachelor's degree, and five percent have a graduate degree. The parents who filled out the enrollment form report on average an annual household income of $31,003 USD and 21 hours worked during a week.[5] On average, parents are 33.3 years old. Sixty-six of parents received the text messages in English, 21 percent in Spanish, and 13 percent in Cantonese. Seventeen percent of all parents participated in text messaging programs in the year before (2016-17) and 1.3 percent did so two years prior (2015-16); all of these parents are in SFUSD.

Table 2 also provides descriptive statistics of parents' answers to math-related parenting questions on the enrollment form. On average, parents agree that they know how to support their children's math development (3.85 on a 4-point scale), report that they count to 20 or higher close

[5] Parents reported household income in six categories: less than $35,000, $35,000 to $49,999, $50,000 to $74,999, $75,000 to $99,999, $100,000 to $149,999, and $150, 000 or more. Income was then transformed into numeric values as the midpoint in each category; the last category was coded as 175,000. Hours were reported in eight categories: I don't work, less than 20 hours, 20-29 hours, 30-39 hours, 40 hours, 41-50 hours, 51-60 hours, more than 60 hours. Again, we assigned the midpoint for each category.

to three to four times per week, and work with them on patterns, use household items to help with math, and play math games around two times per week.

Table 2: Descriptive Statistics

Variables	Mean	Standard Deviation	N
Child Characteristics			
Female	0.495		1331
Age	4.131	0.491	1331
Asian	0.336		1331
Black	0.138		1331
Hispanic	0.392		1331
White	0.065		1331
Other Race/Ethnicity	0.070		1331
Missing Race/Ethnicity Information	0.004		1336
Parent Characteristics			
Less than High School	0.188		756
High School	0.311		756
Some College	0.253		756
Associate's Degree	0.083		756
Bachelor's Degree	0.112		756
Advanced Degree	0.053		756
Missing Education Information	0.434		1336
Age	33.271	6.610	778
Household Income	31002.72	28369.90	698
Hours Worked	20.898	16.977	744
Received Texts in English	0.664		1336
Received Texts in Spanish	0.208		1336
Received Texts in Chinese	0.128		1336
Received Texts in 2015-2016	0.013		1336
Received Texts in 2016-2017	0.170		1336
Parent Baseline Survey Responses on Math Related Items			
Knows How to Build Math Skills	3.845	1.027	756
Counts to 20 or Higher With Child	2.703	0.970	778
Works on Patterns with Child	2.474	1.002	775
Uses Household Objects to Help With Math	2.138	0.975	774
Plays Math Games	2.039	0.957	779

Notes: Parents rated their agreement with a statement that they had knowledge of how to build their child's math skills on a five-point scale (1-Strongly Disagree; 2-Disagree; 3-Neither Agree nor Disagree; 4-Agree; 5-Strongly Agree). Parents rated the frequency of engaging in math activities on a four-point scale (1-Not at All; 2-Once or Twice; 3-Three or Four Times; 4-More Than Four Times).

Estimation

A. Treatment Effects

Given the random assignment, we estimate the effects of the pure math and combination program with the following model:

$$y_{is} = \alpha + \beta_1 \cdot Combo_{is} + \beta_2 \cdot Math_{is} + \delta \cdot X_{is} + \gamma_s + \varepsilon_{is}, \qquad (1)$$

where y_{is} is the outcome of interest of student, i, in randomization block (center by language), s. The outcomes are the overall mean math score as well as the scores for the different assessment components (i.e. rote counting, sorting, etc.). We standardize all scores to have a mean of zero and a standard deviation of one. X_{is} is a vector of student and parent characteristics. This vector includes student age, parent age, household income, hours worked, fall DRDP scores, mean parent responses to parenting questions, indicators for student gender, race/ethnicity, parent education, and parent participation in texting program in previous years, and indicators for missing information.[6] These covariates are listed in the randomization checks found below in Tables 3a and 3b. γ_s is a vector of randomization block fixed effects and ε_{is} is a student level error term. $Combo_{is}$ and $Math_{is}$ indicate that the parent received the combination or pure math text messaging program, respectively. The omitted category is the control group. The estimates of β_1 and β_2 can be interpreted as the causal average treatment effects of the combination and pure math treatment in comparison to the control group. In all regressions, we cluster standard errors on the randomization block level. Appendix Table B.1 presents how the sample, centers, and randomization blocks are distributed across treatment arms and by child gender.

We also probe for heterogeneity of results by gender and skill level.[7] To estimate effects by gender, we fit Equation 1 separately on the subgroup of boys and girls. Within each subgroup we estimate the effect of the combination and pure math programs relative to the control group. We then test whether the estimated effect of each program is equal for boys and girls. Unfortunately, children were not assessed in math at the beginning of the school year, and therefore

[6] Binary covariates with missing information were set to zero and non-binary covariates with missing information were imputed with the district mean. Additionally, we included dummy variables to account for this imputation.
[7] This experiment was pre-registered at Open Science Framework, which included both our main analysis and heterogeneity analyses by gender, race/ethnicity, and skill level. The full pre-registered analysis plan can be accessed at https://osf.io/w6qhr/

we cannot assess effect heterogeneity with respect to baseline skills in math.[8] However, we can investigate how the effects for girls and boys are concentrated along the outcome distribution.

To that end, we estimate quantile regressions for girls and boys separately. Quantile regressions estimate the effect of the program for each quantile of the outcome distribution specified. We use models with and without the baseline covariates in Equation 1. One challenge in comparing quantile effects between girls and boys is that the distributions of math assessment scores differ for girls and boys. A given quantile of the girls' distribution may not correspond to the same quantile of the boys' distribution. To address this challenge, we follow Bitler, Hoynes, and Domina (2014) and Strittmatter (2019) and calculate *translated* quantile effects. Translated quantile effects assign the quantile effects for girls and boys to the same absolute scale of a reference distribution. In this vein, we first calculate the effect at the original quantile for girls and boys. We then find the quantile in the reference distribution that corresponds to the math score at the original quantile and record the quantile effect at the reference quantile. We chose the math score distribution of the control group as the reference distribution.

B. Randomization Checks

Causal identification of the program effects relies on successful randomization. That is, the two treatment groups and the control group do not systematically differ in observed and unobserved characteristics other than being assigned to one of our text messaging programs. To partially test this assumption, we assess covariate balance across treatment arms with the following fixed effects regression model:

[8] Though the DRDP can be used as a measure of baseline math skills, the shortcomings of this assessment do not make it an ideal measure. The assessment asks teachers to rate students on 43 skills via observation, not one-one-assessments. As discussed in the *Potential Teacher Role in Gender Differences* section, teacher perception of skills and third-party assessment of skills do not always align. Further, correlations between domains of the DRDP indicate that teachers do not effectively distinguish between child development domains. Correlations available upon request.

$$X_{is} = \alpha + \beta_1 \cdot Combo_{is} + \beta_2 \cdot Math_{is} + \gamma_s + \varepsilon_{is} \qquad (2)$$

The coefficients of interest are β_1 and β_2, which represent an estimate whether the covariate of interest, X_{is}, is statistically significant between the control group and the combination program group and pure math program group, respectively. If randomization was successful most coefficients should be quantitatively small and statistically insignificant. Tables 3a and 3b show that the groups that received the combination and math programs do not significantly differ from the control group in any of the observed student and parent characteristics. Merely one F-test of the joint significance of both programs (out of 32 tests) is significant at the five percent level. Appendix Tables B.2 through B.5 present covariate balance by gender and once again show that the rate of statistically significant balance is what one would expect by chance. As such, these results provide evidence that randomization was successful and that there are no meaningful differences in observed characteristics. However, our preferred model includes these covariates to increase precision.

C. Attrition

Causal identification could also be jeopardized if there was differential attrition among the three arms of the experiment. Of the 1,842 children recruited into the experiment and randomized to groups, we were unable to assess 452 children due to child absences,[9] 33 children because the teacher indicated that they did not want assessors to test that particular child (e.g., if they were special education), 14 children because the assessor ran out of time before assessing all children or forgot a child, and seven children because of their behavior (e.g., child could not sit still) or because the assessor did not speak the child's language.

[9] Unfortunately, our data does not allow us to distinguish whether students had left the district or were simply absent on the date of the assessment.

Table 3a: Covariate Balance of Student Characteristics

	(1) Combo	(2) Math	F-Test (p-Value)	N
Female	0.015	-0.015	0.747	1336
	(0.039)	(0.039)		
Age	-0.021	0.004	0.394	1336
	(0.019)	(0.020)		
Asian	0.018	0.032	0.375	1336
	(0.023)	(0.023)		
Black	-0.023	-0.012	0.488	1336
	(0.019)	(0.020)		
Hispanic	0.000	-0.013	0.844	1336
	(0.025)	(0.025)		
White	0.015	-0.001	0.626	1336
	(0.018)	(0.014)		
Other	-0.006	-0.001	0.953	1336
	(0.020)	(0.018)		
Missing Race/Ethnicity Information	0.001	-0.003	0.744	1336
	(0.005)	(0.005)		
Fall DRDP				
Approaches To Learning	0.01	0.039	0.83	1298
	(0.068)	(0.070)		
Social and Emotional Development	-0.002	0.011	0.975	1298
	(0.064)	(0.069)		
Language and Literacy Development	0.006	0.041	0.793	1298
	(0.062)	(0.069)		
Cognition, Including Math/Science	-0.007	0.004	0.982	1298
	(0.064)	(0.070)		
Physical Development and Health	0.016	-0.01	0.918	1298
	(0.063)	(0.069)		

Notes: Fall DRDP domain averages activities are standardized to have mean zero and standard deviation one. All models include randomization block fixed effects. Standard errors are clustered at the randomization block level. N = 1,336. * indicates $p < 0.05$

Table 3b: Covariate Balance of Parent Characteristics

	(1) Combo	(2) Math	F-Test (p-Value)	N
Parent Education				
Less Than High School	0.000	-0.002	0.997	1336
	(0.025)	(0.024)		
High School	-0.031	-0.038	0.354	1336
	(0.026)	(0.028)		
Some College	-0.003	0.019	0.676	1336
	(0.025)	(0.027)		
Associate's Degree	0.036	0.022	0.045	1336
	(0.015)	(0.014)		
Bachelor’s Degree	-0.002	0.000	0.996	1336
	(0.019)	(0.020)		
Master's Degree or Higher	-0.009	0.013	0.234	1336
	(0.011)	(0.013)		
Missing Education Information	0.009	-0.013	0.684	1336
	(0.024)	(0.026)		
Parent Age	-0.737	-0.206	0.549	778
	(0.711)	(0.615)		
Parental Income	1365.652	1800.132	0.733	698
	(2824.145)	(2458.565)		
Hours Worked	-0.704	-0.594	0.934	744
	(2.063)	(1.895)		
Texting Language				
English	-0.01	0.006	0.762	1336
	(0.020)	(0.019)		
Spanish	0.008	0.004	0.894	1336
	(0.017)	(0.015)		
Chinese	0.002	-0.01	0.566	1336
	(0.013)	(0.011)		
Received Texts in 2015-2016	0.005	0.003	0.814	1336
	(0.007)	(0.008)		
Received Texts in 2016-2017	0.019	0.019	0.382	1336
	(0.016)	(0.016)		
Average Parent Baseline Reports of Attitudes and Activities				
Knows How to Support Literacy/Math/Behavior	0.008	0.002	0.997	786
	(0.103)	(0.104)		
Frequency of Literacy Related Activities	-0.017	0.037	0.879	789
	(0.094)	(0.098)		
Frequency of Math Related Activities	-0.025	-0.074	0.766	792
	(0.111)	(0.103)		
Frequency of Behavior Related Activities	0.061	-0.032	0.642	791
	(0.108)	(0.106)		

Notes: Parent baseline reports of attitudes and activities are domain averages of standardized items. Averages are standardized to have mean zero and standard deviation one. All models include randomization block fixed effects. Standard errors are clustered at the randomization block level. N = 1,336. * indicates $p < 0.05$

It is unlikely that the text messaging program led students to be absent during the assessment and thus influenced who is part of our estimation sample. However, if that had been the case and these additional students differed on average from students in the control group, the estimated effects of equation (1) would be biased. Therefore, to assess selective attrition, we estimate the following fixed effects model:

$$A_{is} = \alpha + \beta_1 \cdot Combo_{is} + \beta_2 \cdot Math_{is} + \delta \cdot X_{is} + \gamma_s + \varepsilon_{is}, \qquad (3)$$

where A_{is} is an indicator for attrition; it takes the value one when a student was not assessed and is therefore not included in the estimation sample and zero otherwise. The coefficients of interest are once again β_1 and β_2 which now represent an estimate whether the probability of attrition from the sample is statistically significant between the control group and the combination program group and pure math program group, respectively. If the program did not affect who is in our final sample then, once again, each coefficient should be quantitatively small and statistically insignificant. Table 4 shows these coefficients. Neither the combination nor the pure math program led to significantly more sample attrition than the control group. Appendix Table B.6 shows no significant differential attrition by program type in separate boys and girls subsamples.

Table 4: Attrition Balance

	(1)	(2)
	Combination	Pure Math
Not Assessed	0.007	-0.032
	(0.024)	(0.023)

Notes: All models include randomization block fixed effects and a covariates detailed in Tables 3a and 3b. Standard errors are clustered at the randomization block level. N = 1,842. * indicates $p < 0.05$

Results

A. Main Results

Table 5 shows the regression results from equation (1) for the mean math assessment score overall and for girls and boys separately. Overall, neither the combination program nor the pure math program had a significant effect on students' math development. However, the average effect masks effect heterogeneity of the combination program by gender. The combination program appears to have increased girls' assessment scores overall by 0.156 standard deviations (SDs). This coefficient is significant at the ten-percent level. The pure math program, in contrast, had no significant effect on either girls or boys. The difference in impact estimates between boys and girls are statistically significant for the combination program only. These results are consistent with the pilot studies where treatment arms were tested in separate districts.[10]

To further probe the effects of our two programs, Table 6 shows the estimation results for each math assessment subscore. Overall, the effects mirror those on the overall score. While neither the combination nor the pure math program had significant effects overall, we see a clear pattern for girls. The estimated coefficients for all but one of the eight scores are positive, and the coefficients for matching quantities with numerals and reading numerals are significantly different from zero at the five percent level. For boys, the effects of the combination program are mostly negative but smaller in magnitude and insignificant, with the exception of the score for knowing missing numerals in sequences (p<0.1). The pure math program did not lead to any significant effects on girls or boys.

Table 5: Effect of Combination and Pure Math Program on Overall Math Achievement

	Mean Math Score				
	Combination		Pure Math	p-Value (Pure Math vs Combination)	N
All Students	0.000		-0.034	0.569	1336
	(0.058)		(0.056)		
Girls	0.156	+	0.015	0.16	661
	(0.083)		(0.099)		
Boys	-0.115		-0.025	0.324	675
	(0.105)		(0.094)		
p-value (Girls vs. Boys)	.0101		.07179		

Notes: Mean math score is the average of the standardized subscores. The average is standardized to have mean zero and standard deviation one. All regression models include randomization block fixed effects and a full set of covariates. Standard errors are clustered on the randomization block level. + indicates p<0.1.

[10] Table B.7 in the appendix shows the results using alternative model specifications: 1) with no randomization block fixed effects but with covariates and 2) with no randomization block FE and no covariates. The results are unchanged.

Table 6: Effect of Combo and Math Program on Math Assessment Subscores

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All Students (N=1336)			Girls (N=661)			Boys (N=675)		
	Combo	Math	p-Value (Math vs Combo)	Combo	Math	p-Value (Math vs Combo)	Combo	Math	p-Value (Math vs Combo)
Number Concepts	-0.013	0.003	0.802	0.104	0.078	0.81	-0.086	-0.054	0.743
	(0.066)	(0.058)		(0.106)	(0.113)		(0.112)	(0.095)	
Comparing Amounts	0.012	-0.059	0.323	0.111	-0.061	0.157	-0.028	0.039	0.558
	(0.073)	(0.077)		(0.105)	(0.124)		(0.133)	(0.115)	
Sorts Objects	-0.083	-0.068	0.814	-0.025	-0.01	0.88	-0.124	-0.133	0.935
	(0.060)	(0.062)		(0.098)	(0.100)		(0.122)	(0.107)	
Quantities to Numerals	0.078	0.002	0.241	0.198*	0.037	0.095	-0.023	-0.008	0.895
	(0.052)	(0.061)		(0.087)	(0.101)		(0.101)	(0.104)	
Reads Numerals	0.035	0.019	0.816	0.211*	0.107	0.333	-0.102	-0.046	0.658
	(0.061)	(0.065)		(0.092)	(0.117)		(0.106)	(0.117)	
Word Problems	0.059	-0.005	0.347	0.099	-0.073	0.121	0.014	0.099	0.493
	(0.065)	(0.064)		(0.097)	(0.116)		(0.115)	(0.126)	
Rote Counting	-0.01	-0.02	0.873	0.103	0.052	0.589	-0.108	-0.037	0.522
	(0.074)	(0.059)		(0.112)	(0.110)		(0.121)	(0.094)	
Missing Numerals	-0.077	-0.064	0.832	0.089	-0.042	0.234	-0.201+	-0.002	0.071
	(0.068)	(0.067)		(0.099)	(0.097)		(0.121)	(0.120)	

Notes: All values are standardized to have mean zero and standard deviation one. Items of math assessment represent number of tasks correct. Rote Counting corresponds to the highest number counted to. All regression models include randomization block fixed effects and a full set of covariates. Standard errors are clustered on the randomization block level. + indicates p<0.1; * p<0.05

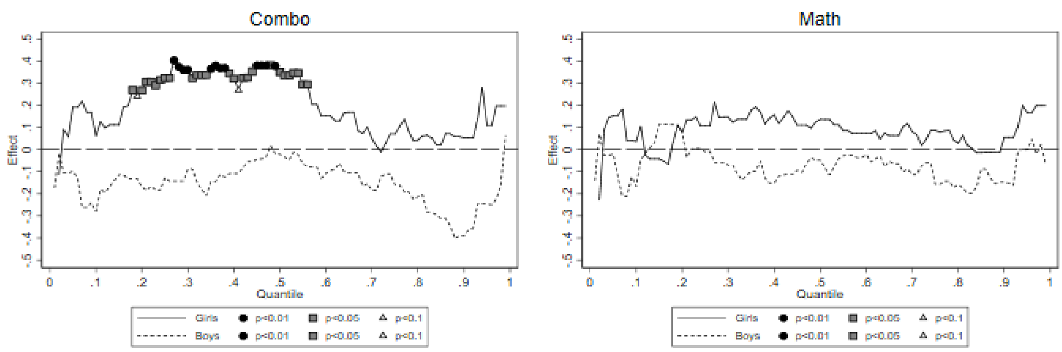
B. Exploring Gender Differences

Math Assessment Differences between Girls and Boys

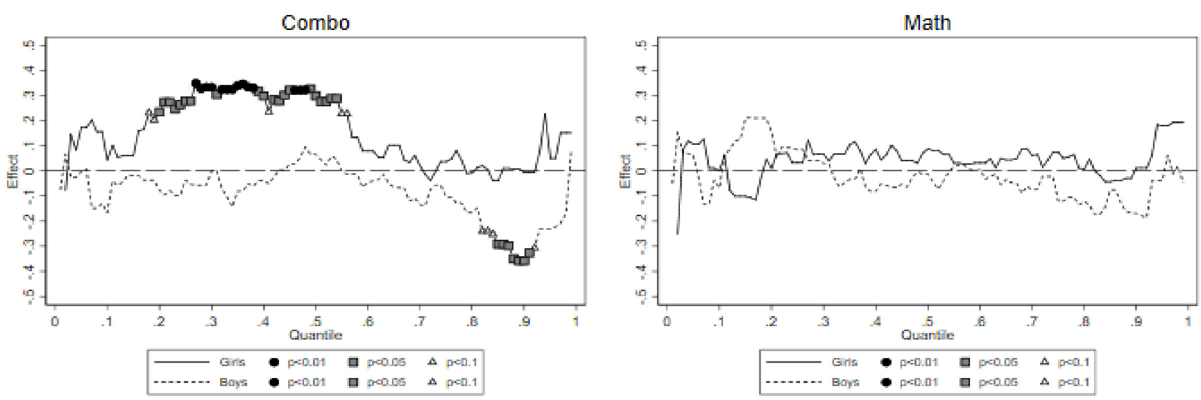
One potential explanation for the observed effect differences between girls and boys is that girls and boys started the year with different baseline skills. Previous studies have repeatedly shown that these programs tend to benefit students with lower baseline skills and therefore help to close achievement gaps (York et al.; 2019; Doss et al., 2019; Cortes et al., in 2021; and Cortes et. al, 2019).

Overall, the translated quantile regression results suggest that the combination program increased girls' math development in the lower half of the outcome distribution. Figure 1 shows the translated quantile effects of the combination and pure math programs on the mean math score for girls and boys. Panel A shows results from the unconditional regression model and Panel B conditional on covariates. Both panels show positive and significant effects of the combination program on girls between the 18-percentile and the 57-percentile of the reference distribution.

Boys in the combination program at the top of the distribution appear to have lower math scores than those in the control group (80-percentile to 92-percentile) and are significant after controlling for covariates. The pure math program does not show any significant quantile effects for girls or boys.



(a) Unconditional Quantile Regression on Overall Math Score



(b) Conditional Quantile Regression on Overall Math Score

Figure 1: Quantile Regressions on Math Outcomes, by Gender

The additional assumption of rank preservation in our quantile results is needed to conclude that the observed results are due to a differential effect of the program by gender by baseline ability. Otherwise, both boys and girls at the bottom of the baseline distribution could have equally benefitted from the program and the overall effect of was driven because girls were simply weaker at baseline. Rank preservation assumes that students weaker in the outcome distribution were also weaker at the baseline distribution and thus the results were only concentrated on girls weaker at the baseline distribution.

This assumption is strong, but can also explain results when investigating whether math scores of girls and boys differ within the treatment and control arms at the end of the school year. Looking at differences within different treatment and control arms allows us to estimate the magnitude of the differences between girls and boys had their parents not received the text messages and thus understand the development of the differences between girls and boys of those parents who received the two treatment programs.

We find that in the control group, girls on average scored lower on the math assessment compared to boys. Table 7 shows math assessment means for boys and mean differences for girls in each treatment group. In the control group, boys outperform girls on average by 0.166 SDs. This is true for almost every task category and significantly so for understanding number concepts, matching quantities with numerals, reading numerals, and identifying missing numerals in sequences. In contrast to the control group, girls that received the combination program outperformed boys that received the combination program on average by 0.191 SDs. This difference is consistently positive across all task categories and significant for comparing different amounts, sorting objects, and rote counting. The differences among students whose parents received the math program are less consistent and mostly insignificant.

Table 7: Gender Differences in Math Outcomes by Experimental Group

	Control (N=440)			Combo program (N=439)			Math program (N=457)	
Outcomes	Dif. Female		Male	Dif. Female		Male	Dif. Female	Male
Mean Math Score	-0.166	+	0.074	0.191	*	-0.083	-0.006	-0.003
Understands Number Concepts	-0.164	+	0.059	0.06		-0.035	0.015	0.018
Compares Different Amounts	0.062		-0.026	0.293	**	-0.116	0.061	-0.065
Sorts Objects	-0.058		0.057	0.194	*	-0.111	0.138	-0.082
Matches Quantities with Numerals	-0.168	+	0.047	0.145		-0.029	-0.014	-0.001

Reads Numerals	-0.291	**	0.121	0.073		-0.032	-0.052		0.042
Solves Word Problems	-0.025		-0.024	0.071		0.01	-0.078		0.028
Rote Counting	-0.084		0.042	0.194	*	-0.092	0.062		-0.036
Knows Missing Numerals in Sequence	-0.224	*	0.147	0.061		-0.068	-0.166	+	0.08

Notes: All values are standardized to have mean zero and standard deviation one. Items of math assessment represent number of tasks correct. Rote Counting corresponds to the highest number counted to. + indicates p<0.1; * p<0.05; and ** p<0.01.

Assuming that the relative ranks of boys and girls did not change in the control group during the year would imply that, absent treatment, girls started the year underperforming boys and the gap persisted throughout the school year. In such a case, our results imply the combination program helped girls catch up to, and surpass, boys by the end of the school year. This finding would be especially plausible if the program was effective for girls weaker at baseline but had little or no effect on boys weaker at baseline. The rank preservation assumption is needed because we do not have baseline achievement data. If ranks were not preserved then these patterns of estimated effects may not be accurate. Regardless, even without this assumption, we see that, on average, girls lagged boys at the end of the year absent the treatment, but surpassed them when their parents received the combination program.

Racial/Ethnic Interactions with Gender Differences

Text message-based programs interact with characteristics of respondents in nuanced ways, as demonstrated by our main finding that the combination program generated greater benefits to girls. However, the results thus far might also suggest that boys could be *negatively* affected by the combination program. Indeed, point estimates from our main regressions were negative, although not statistically significant, and quantile regressions showed significant negative effects for boys at the 80^{th}-90^{th} percentile when including covariates. Thus, we further assess whether boys were systematically hurt from the combination program.

Other than gender, language and cultural differences across race/ethnicity generate heterogeneity that can moderate how families interface with the content of the intervention. Sample sizes are too small to look at results independently for each race/ethnicity, thus we look at samples that remove Hispanic and Asian families respectively, the two largest races/ethnicities in our sample. Comparing the results in these subsamples to the main findings allows us to explore characteristics that may be driving our initial findings.

When Hispanic families are removed from the sample, the trends from the quantile results for boys largely remains unchanged (see Figure B.1 in Appendix B). Positive effects remain for girls in the middle of the outcome distribution and potential negative effects for boys remain. These results suggest Hispanic families are not driving the overall negative pattern for boys. However, Figure 2 shows that when Asian families are removed from the sample, the negative effect for boys is no longer present at any part of the outcome distribution. Further there is a more consistent and positive effect for girls at the approximately the 25th percentile of the outcome distribution and higher. In all, these results indicate that all boys are not systematically negatively affected by the combination program but suggests that families of Asian boys, particularly at the top of the outcome distribution, may be interacting with the text messages differently than other families in our sample. We do not have the data to further disentangle potential explanations, which should be a focus of future research.

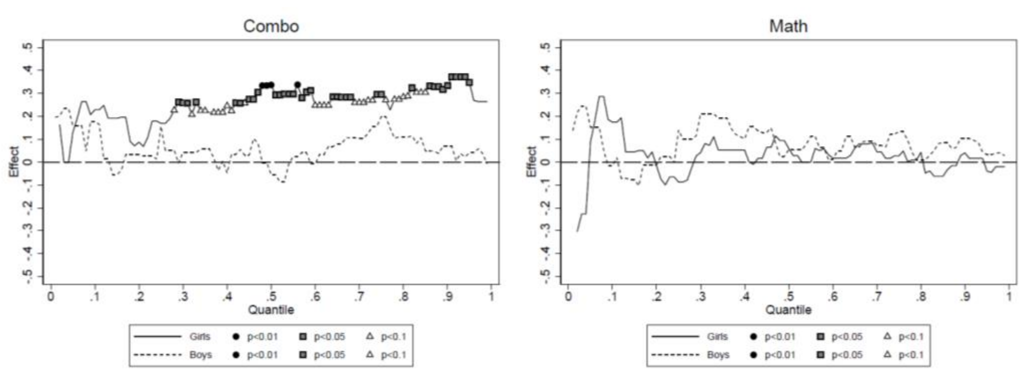


Figure 2: Conditional Quantile Regression on Overall Math Score, Excluding Asian Families

C. Potential Teacher Role in Gender Differences

Teachers' assessment of students' abilities may also contribute to the observed gender differences. Table 8 displays means of baseline characteristics for boys and the mean difference for girls. Teachers assess girls' development higher than boys' even though girls have lower math performance on the standardized assessment. Panel A shows means and mean differences of the fall DRDP teacher assessment subscores.[11] All characteristics are standardized to have a mean of zero and a standard deviation of one. On average, teachers assess the achievement of girls higher than that of boys at baseline. This difference holds for all dimensions of the DRDP and significantly so for four of the five dimensions including for math and science.

This differential teacher assessment has been found elsewhere. A study in Head Start centers found that teachers rated girls as displaying more math interest than boys, though third-party observations of the same children found no differences in interest by gender (Fisher et al., 2012). Similarly, elementary school teachers have been shown to attribute success in early math for boys to inherent ability and for girls to both ability and effort (Gunderson et al., 2012).

Given that teachers often target supports to lower performing students (Rochmes, Penner and Loeb, 2019), these results may suggest that teachers would focus more on boys' development. If teachers focus more on boys when building math skills in class, then the extra in class work would lead boys in both the treatment and control group to learn more of the skills covered in the texting program. The extra knowledge gained in school could mitigate any benefits of in-home work on the same skills and the program would be less effective for boys. We do not, however, have a compelling method for assessing this effect.

[11] Girls and boys did not significantly differ on other demographic characteristics. We therefore did not include these characteristics in Table 8.

Table 8: Differences of Baseline Characteristics Between Girls and Boys

	Dif. Female		Male	N
Panel A: Student Fall DRDP Teacher Assessment				
Approaches to Learning	0.213	***	-0.11	1298
Social and Emotional Development	0.195	***	-0.101	1298
Language & Literacy Development	0.174	**	-0.092	1298
Cognition, Including Math/Science	0.104	+	-0.057	1298
Physical Development & Health	0.078		-0.044	1298
Panel B: Parent Characteristics				
Knows What They Can Do to				
Help Child Develop Literacy Skills	0.067		-0.034	785
Help Child Develop Math Skills	0.146	*	-0.074	750
Improve Child's Behavior	0.138	+	-0.07	774
Frequency of Parent/Child Activities				
Asked Questions about Books	0.154	*	-0.079	784
Practiced Rhyming	0.069		-0.035	772
Introduced New Words	0.113		-0.058	770
Worked on Literacy Skills During Fam. Activities	0.021		-0.011	770
Counted to 20 or higher	0.012		-0.006	778
Worked on Patterns	0.11		-0.056	775
Used Household Objects to Learn Math Skills	0.179	*	-0.092	774
Played Math Learning Games	0.034		-0.017	779
Talked about Feelings	0.16	*	-0.082	782
Gave Choices	-0.045		0.023	783
Praised Effort	0.033		-0.017	765

Notes: + indicates p<0.1; * p<0.05; ** p<0.01, and *** p<0.001.

D. Potential Parent and Child Role in Gender Differences

The differential results may also be a consequence of how participants interacted with the texting program. There are two possible ways in which interactions between the texting program can differ based on the gender of the child: (1) the parent can operationalize the activities differently, either by differentially choosing whether to attempt the activity or by interacting with girls and boys in different ways, and (2) the child can react differently to same activity prompt. Research suggests that either of these mechanisms are plausible.

In support of the first hypothesis, researchers have found that parents perceive the academic ability of boys and girls differently, especially in math. Mothers of boys, for example, have been shown to believe that their child has more talent in math and needs to try less hard to achieve success in math (Gunderson et al., 2012). One possibility, then, is that parents of boys were less likely to take up the math activities, as seen in our results, because they were overconfident in boys' ability to do math or thought they needed less practice. Indeed, research provides evidence that children internalize messages that math is for boys, but this over-confidence may not always be beneficial for boys. Confidence is negatively correlated to math achievement for boys in second grade, but positively correlated to achievement for girls, suggesting that girls may have a more accurate belief regarding their math ability, while boys may be suffering from over confidence (Carr et al., 2008).

When it comes to perceptions of interest in academics, parents perceive girls to be more interested than boys – whether or not this reflected a child's actual interest. A study of five-year-olds in Head Start found that both parents and teachers rated girls as more interested in literacy, though no differences were found in child self-reported interest in literacy (Baroody and Diamond, 2013). In the context of this intervention, parents of girls may be more likely to engage with the activities if they are responding to a greater perceived interest from their child, regardless of whether there is a differential interest in reality. Panel B of Table 8 shows parents' responses to the enrollment form questions about parent-child engagement.[12] Parents of girls report a significantly higher confidence on how to support math and behavioral development than parents of boys. Parents of girls also report a higher frequency of parent-child activities such as asking questions about books, using household items to learn math skills, and talking about feelings.

Even when parents of boys and girls attempt the activities at equal rates, boys and girls may react differently. Early differences in skills, such as regulating emotions and impulses,

[12] Demographic baseline characteristics did not differ between girls and boys or between parents of girls and boys.

between boys and girls may lead to differences in their interactions with the activities. Studies often find that girls outperform boys on these types of skills. Gender differences have been found in children as young as three years old (Ponitz et al., 2008), as well as in elementary school (Matthews, Ponitz, and Morrison, 2009; Else-Quest et al., 2006). They have also been found to affect parent-child interactions where girls are more likely to exhibit a sincere commitment to following maternal requests, such as requests to do challenging and tedious tasks (Kochanska, Coy, and Murray, 2001). Further, these types of skills often predict performance on early literacy and math assessments (Matthews, Ponitz, and Morrison, 2009; Ponitz et al., 2008).

Thus, the concentration of the effects on girls might, in part, reflect a greater ability of girls to regulate their emotions and impulses and follow the requests of their parents. Such a dynamic would increase the probability that the child performed the activity successfully. Greater success in completing the activities could then create a feedback loop where it breeds a greater adherence to the program and future activities. Of course, the greater ability of girls to regulate their emotions may interact with parents' gendered views of their child's interest in academics and math ability to amplify the effects of the program on girls.

Conclusion

This study demonstrates that while common behavioral barriers exist to improving multiple academic domains in the home, behavioral interventions interact with social factors and participants in ways that can produce heterogeneities in effects. In this study, we find that a text messaging program for parents, based on behavioral economics principles and early childhood learning standards, can improve the early mathematics development of prekindergarteners. However, we find that a program that includes literacy and social-emotional skill goals for children is more effective at raising math achievement than a program that is focused purely on mathematics. Further, we find that the benefits are concentrated in girls. A retrospective analysis of the combination program fielded in 2014-2015 in SFUSD and part of York, Loeb, and Doss' (2019) original set of experiments did not find such a gender difference on literacy outcomes. In that study, the effects were concentrated on students that scored below the median of the baseline skill and the evidence suggests that both genders benefited among this subgroup.[13] Thus, this differential effect may be particular to math outcomes.

The ability of these programs to change adult behaviors implies that parents may be underinvesting in their children's early math development. The same behavioral barriers – such as those stemming from high cognitive load and limited attention – that reduce parent investment in building their children's literacy skills may be amplified in math given that many parents have math anxiety, that the importance of early math development has only more recently been documented and emphasized, and that parents more often see math development as the responsibility of schools (Sonnenschein Metzger, and Thompson, 2016; Eden, Heine, and Jacobs, 2013; Gunderson and Levine, 2011; Hembree, 1990).

The greater effectiveness of the combination program offers lessons about in-home math learning processes that may inform parenting support beyond text messaging programs. Programs aiming to overcome informational and behavioral barriers by solely providing math-related information and activities may not be effective, perhaps because parents are not as primed to work on math with their children or because parents may have anxiety associated with their own math experiences. Rather, cycling through literacy, math, and SEL topics may keep the parents more engaged and help them overcome the extra barriers that math content creates. If parents struggle on one domain, they may find success in another. Moreover, these domains are not mutually

[13] Results available upon request.

exclusive (Butterworth 2005; Graziano et al. 2007; Sarama et al. 2012; Morris et al. 2013) and may be complementary even at this early age (Purpura, 2011). As such, combination programs may help ease parents into supporting their children's math development.

Our analyses show that the positive effects of the combination program are concentrated on girls. In fact, quantile regression results indicate that boys in the 80th to 90th percentile scored worse than the control group after receiving the combination program. Further analysis indicates that this result is driven by Asian boys. Parents face common behavioral barriers – such as incomplete information, limited attention, cognitive complexity, and discounting the future – regardless child's gender or race. However, interactions between the child and parent do often differ by gender and cause parents to operationalize program differently. We hypothesize that parents may differentially operationalize the activities due to differences in their perceptions of the academic ability and interest of girls and boys (Fisher et al., 2012; Gunderson et al., 2012; Barrody and Diamond, 2013). Children may also differently react to the prompt to engage in the activities, possibly due to differences in executive functions such as self-regulation (Matthews, Ponitz, and Morrison, 2009; Ponitz et al., 2008; Else-Quest et al., 2006). Our data do not allow us to probe the more nuanced intersection with Asian boys, but results may stem from cultural interactions with the program and its context. Recent research indicates that the cultural relevance of educational interventions can be enormously impactful (Bonilla, Dee, and Penner, 2021; Dee and Penner, 2017).

These nuanced results provide some evidence that, as researchers conceptualize and operationalize behavioral interventions, they may benefit from going beyond the application of broad behavioral economics principles and attempt to understand and incorporate heterogeneity in behavioral barriers and response to those behavioral barriers. Social factors such as gendered beliefs of children’s abilities and attitudes, gendered reactions of children to activities, and the alignment of the skill of the recipient and the proposed task may all have dramatic effects on the efficacy of the program. Future research could attempt to understand these contextual factors. Specifically, in our context, more research is needed to understand why only girls benefited from this intervention, why a segment of higher performing boys (particularly Asian boys) did not score as well, and how the math development of boys can be supported. Doss et al.’s (2019) study showed that matching the skill of the child to the difficulty of texts can improve the effect of the program, particularly for higher performing children. This study reveals that gender may be another important dimension to tailor these programs.

Overall, this intervention was successful in improving the mathematics outcomes for a portion of the sample. To the best of our knowledge, we are the first to support early childhood math development with a text message-based program. The study provides evidence that sustained light-touch interventions can help parents change behavior that meaningfully impacts child learning. The intervention also retains many compelling features of texting interventions, namely the ease of scalability and the cost effectiveness of the learning gains realized.

Acknowledgements and Funding

We thank Petra Thiemann for valuable feedback on this paper. We thank Erika Byun and Kristine Gaffaney, without whom we would not have pulled off the study.

This study was supported by the Heising-Simons Foundation, San Francisco, CA. The broader project was supported by the Arnold Foundation, Houston, TX; the Evelyn and Walter Haas, Jr. Fund, San Francisco, CA; the David and Lucile Packard Foundation, Los Altos, CA; the Silver Giving Foundation, San Francisco, CA; the Smith Richardson Foundation, Westport, CA; and the Spencer Foundation, Chicago, IL.

References

Ashcraft, M. H., and Moore, A. M. (2009). Mathematics anxiety and the affective drop in performance. *Journal of Psychoeducational Assessment, 27*(3), 197-205.

Avery, C., and Kane, T.J. (2004). Student perceptions of college opportunities. The Boston COACH Program. In Caroline Hoxby (ed.) *College choices: The economics of where to go, when to go, and how to pay for it*, pp. 355-394, Chicago: The University of Chicago Press.

Baroody, A.E., and Diamond, K.E. (2013). Measures of preschool children's interest and engagement in literacy activities: Examining gender differences and construct dimensions. *Early Childhood Research Quarterly, 28*(2), 291-301.

Bergman, P. (2015). Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment. *CESifo Working Paper Series, 5391*. Retrieved from https://ssrn.com/abstract=2622034.

Bergman, P., and Chan E.W. (2017). Leveraging Parents through Technology: The Impact of High-Frequency Information on Student Achievement. *CESifo*. Retrieved from https://www.cesifo-group.de/dms/ifodoc/docs/Akad_Conf/CFP_CONF/CFP_CONF_2017/ee17-Hanushek/Papers/ee17_Bergman.pdf

Berkowitz, T., Schaeffer, M.W., Maloney, E.A., Peterson, L., Gregor, C., Levine, S.C., and Beilock, S.L. (2015). Math at home adds up to achievement in school. *Science, 350*(6257), 196-198

Bittler, M., Hoynes, H., and Domina, T. (2014). Experimental Evidence on Distributional Effects of Head Start. NBER Working Paper No. 20434. Retrieved from https://www.nber.org/papers/w20434.

Bonilla, S., Dee, T.S., and Penner, E.K. (2021). Ethnic students increases longer-run academic engagement and attainment. *Proceedings of the National Academy of Sciences, 118*(37).

Butterworth, B. (2005). The development of arithmetic abilities. *Journal Child Psychology and Psychiatry, 46*(1), pp. 3-18.

Bradley, R.H., Corwyn, R.F., Burchinal, M., McAdoo, H.P., and Coll, C.G. (2001). The home environments of children in the United States part II: Relations with behavioral development through age thirteen. *Child Development, 72*(6), pp. 1868-1886.

Cannon, J., Ginsburg, H.P. (2008). "Doing the Math": Maternal beliefs about early mathematics versus language learning. *Early Education and Development, 19(*2), 238-260.

Califonia Department of Education. (2008). California preschool learning foundations, volume 1. Sacramento, CA. Retrieved, December 15, 2015 from http://www.cde.ca.gov/sp/cd/re/documents/preschoollf.pdf

Carr, M., Steiner, H.H., Kyser, B., and Biddlecomb, B. (2008). A comparison of predictors of early emerging gender differences in math competency. *Learning and Individual Differences, 18,* 61-75.

Castleman, B.L., and Page L.C. (2015). Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates? *Journal of Economic Behavior & Organization*, *115,* pp. 144-160.

Castleman, B.L., and Page L.C. (2016). Freshman Year Financial Aid Nudges. *Journal of Human Resources*, *51*(2), pp. 389-415.

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

Clements, D. (2004). Preface. *Engaging young children in mathematics: Standards for early childhood mathematics education.* Eds. Clements, D.H. and Samara, J. Mahwah, NJ: Lawrence Earlbaum Associates

Cohen, G. L., Sherman, D. K. (2014). The psychology of change: Self-affirmation and social psychological intervention. *Annual Review of Psychology, 65,* pp. 333-371.

Committee on Stem Education. (2018). Charting a course of success: America's strategy for STEM education. *National Science and Technology Council.* Retrieved from: https://www.whitehouse.gov/wp-content/uploads/2018/12/STEM-Education-Strategic-Plan-2018.pdf

Cortes, K., Fricke, H., Loeb, S., Song, D, and York, B. (2021). Too little or too much? Actionable advice in an early-childhood text messaging experiment. *Education Finance and Policy, 16*(2), pp. 209-232.

Cortes, K., Fricke, H., Loeb, S., Song, D, and York, B. (2019). When Behavioral Barriers are Too High or Low - How Timing Matters for Parenting Interventions. NBER Working Paper No. 25964. Retrieved from https://www.nber.org/papers/w25964.

Cvencek, D., Meltzoff, A.N., and Greenwalk, A.G. (2011). Math-gender stereotypes in elementary school children. *Child Development, 82*(3), 766-779

Dee, T.S. and Penner, E.K. (2017). The causal effects of cultural relevance: Evidence from an ethnic studies curriculum. *American Educational Research Journal*, *54*(1), pp. 127-166.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, *47*(2), pp. 315-372.

Doss, C., Fahle, E., Loeb, S., and York, B. (2019). Supporting parenting through differentiated and personalized text messaging: Testing effects on learning during kindergarten. *Journal of Human Resources*, 54(3), pp. 537-566.

Duncan GJ, Ziol-Guest KM, Kalil A. Early childhood poverty and adult attainment, behavior, and health. *Child Development.* 2010; 81: 306–325.

Eden, C., Heine, A., and Jacobs, A.M. (2013). Mathematics anxiety and its development in the course of formal schooling – A review. *Psychology, 4*(6), 27-35.

Else-Quest, N.M., Hyde, J.S., Goldsmith, H.H., and Van Hulle, C.A. (2006). Gender differences in temperament: A meta-analysis. *Psychological Bulletin*, *132*(1), 33-72.

Fisher, P.H., Dobbs-Oats, J., Doctoroff, G.L., and Arnold, D.H. (2012). Early math interest and the development of math skills. *Journal of Educational Psychology*, *104*(3), 673-681.

Fricke, H., Grogger, J., and Steinmayr, A. (2018). Exposure to academic fields and major choice. *Economics of Education Review, 64*(June), pp. 199-213

Fricke, H., Kalogrides, D., and Loeb, S. (2018). It's too annoying: Who drops out of educational text messaging programs and why. *Economics Letters*, *173*, pp. 39-43.

Graziano, P. A., Reavis, R.D., Keane, S.P., and Calkins, S.D. (2007). The role of emotion regulation in children's early academic success. *Journal of School Psychology*, *45*(1), pp. 3-19.

Grodsky, E., and Jones, M.T. (2007). Real and imagined barriers to college entry: *Perceptions of cost. Social Science Research*, *36*(2), pp. 745-766.

Groot, B., Sanders, M., Rogers, T., and Bloomenthal, E. (2017). I get by with a little help from my friends: Two field experiments on social support and attendance in further education colleges in the UK. *The Behavioral Insights Team*. Retrieved from: http://38r8om2xjhhl25mw24492dir.wpengine.netdna-cdn.com/wp-content/uploads/2017/04/Study-Supporter-WP_April-2017.pdf.

Gunderson, S.A. and Levine, S.C. (2011). Some types of parent number talk count more than others: relations between between parents' input and children's cardinal-number knowledge. *Developmental Science*, *14*(5), 1021-1032.

Gunderson, E.A., Ramirez.G., Levine, S.C., and Beilock, S.L. (2012). The role of parents and teachers in the development of gender-related math Attitudes. *Sex Roles*, *66*, 153-166

Hall, C. C., Zhao, J., and Shafir, E. (2014). Self-affirmation among the poor: cognitive and behavioral implications. *Psychological Science*, *25*(2), pp. 619-625.

Hastings, J.S., and Weinstein, J.M. (2008). Information, school choice, and academic achievement: Evidence from two experiments. *Quarterly Journal of Economics*, 123(4), pp. 1373-1414.

Heckman, J.J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, *312*(5782), pp. 1900-1902.

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education, 21*(1), 33-46.

Iyengar, S., and Lepper, M. (2000). When choice is demotivating: Can one desire too much of a good thing? *Personality & Social Psychology, 79*(6), pp. 995-1006.

Jordan, N.C., Kaplan, D., Ramineni, C., and Locuniak, M.N. (2009). Early Math Matters: Kindergarten Number Competence and Later Mathematics Outcomes. *Developmental Psychology, 45*(3), 850-867

Karlan, D., McConnell, M., Mullainathan, S., and Zinman, J. (2016). Getting to the Top of Mind: How Reminders Increase Saving. *Management Science,* 62(12), pp. 3393-3672.

Kochanska, G., Coy, K.C., and Murray, K.T. (2003). The development of self-regulation in the first four years of life. *Child Development, 72*(4), 1091-1111.

Kraft, M.A., and Rogers, T. (2015). The underutilized potential of teacher-to-parent communication: Evidence from a field experiment. *Economics of Education Review, 47*, pp. 49-63.

Levine, S.C., Suriyakham, L.W., Rowe, M.L., Huttenlocher, J., and Gunderson, E.A. (2010). What counts in the development of young children's number knowledge? *Developmental Psychology, 46*(5), 1309-1319.

Lonigan, C.J., and Shanahan, T. (2009). Developing early literacy: Report of the National Early Literacy Panel. *National Institute for Literacy*. Retrieved, December 15, 2005 from https://eric.ed.gov/?id=ED508381.

Luttenberger, S., Wimmer, S., and Paechter, M. (2018). Spotlight on Math Anxiety. *Psychology Research and Behavior Management, 11*, 311-322.

Maloney, E.A., Ramirez, G., Gunderson, E.A., Levine, S., and Beilock, S. (2015). Intergenerational effects of parents' math anxiety on children's math achievement and anxiety. *Psychological Science, 26*(9), 1480-1488

Mani, A., Mullainathan, S., Shafir, E., and Zhao, J. (2013). Poverty impedes cognitive function. *Science*, *341*(6149), pp. 976-980.

Manolitsis, G., Georgiou, G.K., and Tziraki, N. (2013). Examining the effects of home literacy and numeracy environment on early reading and math acquisition. *Early Childhood Research Quarterly, 28*, 692-703

Matthews, J.S., Ponitz, C.C., and Morrison, F.J. (2009). Early gender differences in self-regulation and academic achievement. *Journal of Educational Psychology, 101*(3), 689-704

Mendelsohn, Alan L., Leora N. Mogilner, Benard P. Dreyer, Joel A. Forman, Stacey C. Weinstein, Monica Broderick, Karyn J. Cheng, Tamara Magloire, Taska Moore, and Camille

Napier. 2001. “The Impact of a Clinic-Based Literacy Intervention on Language Development In Inner-City Preschool Children.” *Pediatrics 107(*1):130-134.

Morris, P., Millenky, M., Raver, C.C., and Jones, S.M. (2013). Does a preschool social and emotional learning intervention pay off for classroom instruction and children’s behavior and academic skills? Evidence from the Foundations of Learning project. *Early Education and Development, 24*(7)*,* pp. 1020-1042.

Mullainathan, S., and Thaler, R.H. (2000). Behavioral Economics. *NBER Working Paper No. w7948*. Retrieved December 15, 2015 from http://www.nber.org/papers/w7948

Nobles, W.P. (2018). Chelsea Clinton debuts “Talking is Teaching” campaign in New Orleans. *Nola.com.* Retrieved from: https://www.nola.com/news/education/article_1ade926f-e25a-5c9e-a2cb-731f92f6d8de.html

Page, L.C., Castleman, B., and Meyer K. (2016). Customized nudging to improve FAFSA completion and income verification. Working Paper. Retrieved from https://ssrn.com/abstract=2854345

Ponitz, C.E.C., McClelland, M.M., Jewkes, A.M., Connor, C.M., Farris, C.L., and Morrison, F.J. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Earlyy Childhood Research Quarterly, 23*(2), 141-158.

Pruden, S.M., Levine, S.C., and Huttenlocher, J. (2011). Childrens’ spatial thinking: does talk about the spatial word matter?. *Developmental Science*, *14*(6), 1417-1430

Pungello EP, Kainz K, Burchinal M, et al. Early educational intervention, early cumulative risk, and the early home environment as predic- tors of young adult outcomes within a high-risk sample. *Child Development.* 2010; 81: 410-426.

Purpura, D.J., Hume, L.E., Sims, D M., and Lonigan, C.J. (2011). Early literacy and early numeracy: The value of including early literacy skills in the prediction of numeracy development. *Journal of Experimental Child Psychology*, *110*(4), pp. 647-658.

Reese, E., Sparks, A., and Leyva, L. (2010). A review of parent interventions for preschool children’s language and emergent literacy. *Journal of Early Childhood Literacy*, *10*(1), pp. 97-117.

Rittle-Johnson, B., Fyfe, E.R., Hofner, K.G., and Farra, D.C. (2017). Early math trajectories: Low-income children’s mathematics knowledge from ages 4-11. *Child Development, 88*(5), 1727-1742.

Robinson, C. D., Lee, M. G., Dearing, E., and Rogers, T. (2018). Reducing student absenteeism in the early grades by targeting parental beliefs. *American Educational Research Journal*. Retrieved from http://journals.sagepub.com/toc/aer/0/0.

Rochmes, J., Penner, E., and Loeb, S. (2019). “Educators as Equity Warriors.” Working Paper.

Rogers, T., and Feller, A. (2018). Reducing student absences at scale by targeting parents' misbeliefs. *Nature: Human Behavior, 2*(2018), pp. 335-342.

Sarama, J., Lange, A.A., Clements, D.H., and Wolfe, C.B. (2012). The impacts of an early mathematics curriculum on oral language and literacy. *Early Childhood Research Quarterly*, *27*(3), pp. 489-502.

Shah, A., Mullainathan, S., and Shafir, E. (2012). The Consequences of Having Too Little." *Science, 338*(6107), pp. 682-685.

Smythe-Leistico, K., and Page, L. C. (2018). Connect-text: leveraging text-message communication to mitigate chronic absenteeism and improve parental engagement in the earliest years of schooling. *Journal of Education for Students Placed at Risk*, *23*(1-2), pp. 139-152.

Sonnenschein, S., Galindo, C., Metzger, S.R., Thompson, J.A., Huang, H.C., and Lewis, H. (2012). Parents' beliefs about children's math development and children's participating in math activities. *Child Development Research, 2012,* 1-13.

Sonnenschein, S., Metzger, S.R., and Thompson, J.A. (2016). Low-income parents' socialization of their preschoolers' early reading and math skills. *Research in Human Development, 13*(3), 207-224

Strittmatter, A. (2019). Heterogeneous earnings effects of the job corps by gender: A translated quantile approach. *Labour Economics*, *61*, 101760.

Sweeney, A. M., and Moyer, A. (2014). Self-affirmation and responses to health messages: A meta-analysis on intentions and behavior. *Health Psychology*, *34*(2), pp. 149-159.

Valant, J., and Loeb, S. (2014). Information, choice, and decision-making: Field experiments with adult and student school choosers. Working Paper.

Weixler, L., Valant, J., Bassok, D., Doromal, J.B., and Gerry, A. (2020). Helping parents navigate the early childhood education enrollment process: Experimental evidence from New Orleans. *Educational Evaluation and Policy Analysis, 42*(3), pp. 307-330.

York, B.N., Loeb, S., and Doss, C. (2019). One Step at a Time: The Effects of an Early Literacy Text Messaging Program for Parents of Preschoolers. *Journal of Human Resources,* 54(3), pp. 567-603.

Appendix A: 2015-2016 Pilot Math Texting Program

Recruitment

The pilot math texting program was instituted with three partners: the San Francisco Unified School District (SFUSD), the Institute for Human and Social Development (IHSD; also known as Head Start) in San Mateo county, and Jumpstart of Northern California. Jumpstart is a non-profit organization that provides academic services to three and four-year-old children. During the summer of 2015, we worked with each partner to recruit parents of three- and four-year-old children into the study through each organization's enrollment process. Interested parents completed a consent form and provided their cell phone number and basic demographic information. Parents in SFUSD completed an additional intake form that asked judgements about their child's academic abilities and the frequency with which they engaged in academically oriented activities with their children. In total 1,186 children were recruited into the study – 434 from IHSD, 238 from Jumpstart, and 494 from SFUSD.

Pilot Study Design

Families in all three sites were randomly assigned to receive either a 32-week text messaging curriculum or *placebo* texts. Placebo texts contained programmatic information, such as information about deadlines and vaccination requirements, and did not contain information pertaining to any academic subject. One placebo text was sent every two weeks. Treatment parents in IHSD and Jumpstart sites received text messages that addressed only math skills. Treatment parents in SFUSD received text messages that address literacy, math, and social-emotional skills. All texting programs started in October 2015 and ended in May 2016. Participants were able to choose text messages in either English, Spanish, or Chinese. Randomization was blocked at the school site.

Math Outcome

Our main math outcome was the Tools for Early Assessment in Math (TEAM) short form, a validated instrument of math knowledge in prekindergarten (Weiland et al., 2012). The TEAM is a pull-out, one-on-one assessment that takes approximately 20 minutes. The TEAM covers the following topics: counting, comparing quantities in groups, subitizing, matching numerals to sets, addition and subtraction, shapes, patterns, measurement, and comparing weights.

We trained 39 assessors to administer the assessment to children in the study (both treatment and control) in all three program sites. The bulk of the assessment occurred during April and May 2016, with a few assessments occurring during that June. Of the 1,186 children in the study, we were able to collect math outcomes on 941 children. The remaining 245 children were either absent on the day of the data collection or were unable to successfully complete any questions. The final sample in SFUSD, IHSD, and Jumpstart is 418, 310, and 213, respectively.

Estimation

Given the random assignment, we estimate the effects of the texting curricula with the following model:

$$y_{is} = \alpha + \beta_1 \cdot Treat_{is} + \delta \cdot X_{is} + \gamma_s + \varepsilon_{is}, \qquad (1)$$

where y_{is} is the TEAM outcome of student, i, in randomization block (center), s. We first standardized (within program) each item of the TEAM assessment. We then average the standardized score, and finally standardizing the average. X_{is} is a vector of student and parent characteristics. For the IHSD and Jumpstart program this vector includes student age and indicators for student gender, home language, texting language, and being in the Jumpstart program.[14] The vector of covariates in the SFUSD sample contains parent age, factors of questions on home reading and math habits and child academic ability from the intake question, indicators for the child's race, parent gender, and whether the parent's highest level of education is high school or less, as well as the covariates in the IHSD and Jumpstart sample. γ_s is a vector of randomization block fixed effects and ε_{is} is a student level error term. $Treat_{is}$ is an indicator that the parent received a texting curriculum, either the pure math curriculum in IHSD and Jumpstart or the combination program in SFUSD. Further, we check covariate balance and differential attrition in the same way as in our main analysis. Tables A1 and A2 show that all covariates were well balanced between treatment and control groups and that attrition was also well balanced between treatment and control groups, respectively.

Table A.1: Covariate Balance on 2015-2016 Math and Combination Texting Programs

	Math Only	Combination
Covariate	Jump Start and San Mateo Head Start	SFUSD
Home Language English	0.006	0.038
	(0.036)	(0.044)
Home Language Spanish	0.014	0.015
	(0.032)	(0.046)
Home Language Chinese	-0.004	-0.013
	(0.015)	(0.032)
Home Language Other	-0.021	-0.045
	(0.015)	(0.043)
Text Language English	-0.001	0.030
	(0.031)	(0.052)
Text Language Spanish	-0.006	0.001
	(0.032)	(0.039)
Text Language Chinese	-0.007	-0.031
	(0.008)	(0.034)
Child Age	-0.009	-0.003
	(0.036)	(0.017)
Child Female	-0.011	-0.025
	(0.036_	(0.043)
Child Race White		-0.004
		(0.027)
Child Race Black		0.047
		(0.030)
Child Race Hispanic		0.012

[14] Binary covariates with missing information were set to zero and non-binary covariates with missing information were imputed with the program mean. Additionally, we included dummy variables to account for this imputation.

	(0.031)
Child Race Asian	-0.054
	(0.032)
Child Race Other	-0.001
	(0.025)
Parent Age	0.800
	(0.721)
Parent Female	0.017
	(0.037)
Parent Education High School or Less	-0.035
	(0.053)
Factor of Parent Reading Questions	0.021
	(0.085)
Factor of Parent Math Questions	0.113
	(0.090)
Factor of Child Questions	-0.077
	(0.086)

Notes: Mean math score is the average of the standardized items. The average is standardized to have mean zero and standard deviation one. All regression models include randomization block fixed effects. Standard errors are clustered on the randomization block. N= 523 in the IHSD and Jumpstart sample and N= 418 in the SFUSD sample. * indicates $p<0.05$.

Table A.2: Attrition on 2015-2016 Math and Combination Texting Programs

	Math Only	Combination
	Jump Start and San Mateo Head Start	SFUSD
Attrition	0.036	-0.032
	(0.031)	(0.029)

Notes: All regression models include randomization block fixed effects and covariates listed in Table A1. Standard errors are clustered on the randomization block. N = 692 in IHSD and Jumpstart samples and N = 494 in the SFUSD sample. * indicates p<0.05.

Appendix B: Sample Details and Specification Checks

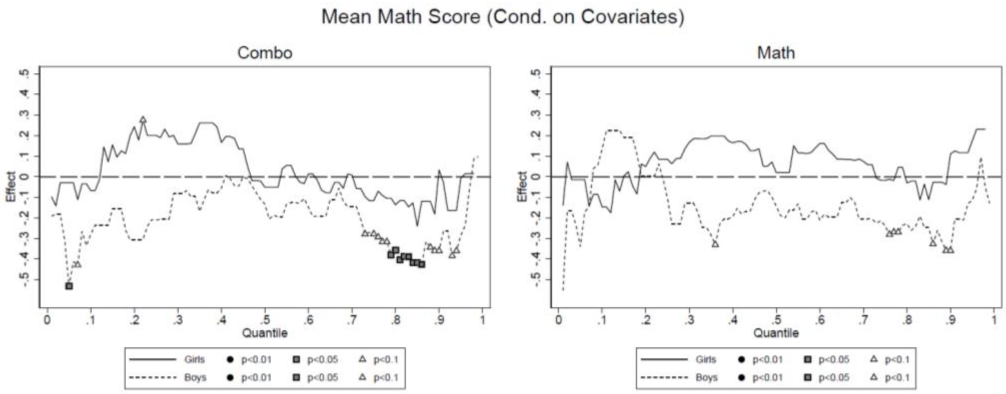


Figure B.1: Conditional Quantile Regression on Overall Math Score, Excluding Hispanic Families

Table B.1: Sample by Randomization Group

	Original Data			Analytic Sample		
	Control	Combination	Pure Math	Control	Combination	Pure Math
Total Sample	612	615	615	440	439	457
Boys	312	316	317	224	216	235
Girls	300	299	298	216	223	222
Sites	79	81	80	68	69	71
Randomization Blocks	210	210	211	173	175	178

Notes: Total sample recruited is 1,842 of which 945 are boys and 897 are girls. Total analytic sample is 1,336 of which 675 are boys and 661 are girls. Texting language-by-site constitute randomization blocks.

Table B.2 Covariate Balance of Student Characteristics, Sample of Boys

	(1)	(2)	(3)	(4)
	Combo	Math	F-Test (p-Value)	N
Age	-0.006	0.008	0.345	675
	(0.007)	(0.010)		
Has 3 Year Old Sibling	0.019	0.001	0.19	675
	(0.013)	(0.015)		
Has Twin	-0.023	0.007	0.058	675
	(0.013)	(0.016)		
Asian	0.082	0.053	0.122	675
	(0.042)	(0.040)		
Black	-0.076*	-0.012	0.016	675
	(0.030)	(0.020)		
Hispanic	0.03	-0.011	0.622	675
	(0.043)	(0.047)		
White	0.017	0.022	0.521	675
	(0.028)	(0.020)		
Other	-0.037	-0.042	0.466	675
	(0.037)	(0.034)		
Missing Race/Ethnicity Information	-0.011	-0.005	0.311	675
	(0.009)	(0.012)		
Fall DRDP				
Approaches To Learning	0.013	0.071	0.755	654
	(0.115)	(0.110)		
Social and Emotional Development	-0.091	-0.015	0.639	654
	(0.111)	(0.103)		
Language and Literacy Development	-0.091	0.021	0.439	654
	(0.106)	(0.098)		
Cognitive Development	-0.146	-0.069	0.432	654
	(0.112)	(0.103)		
Physical Development and Health	-0.074	-0.021	0.744	654
	(0.103)	(0.108)		

Notes: Fall DRDP domain averages activities are standardized to have mean zero and standard deviation one. All models include randomization block fixed effects. Standard errors are clustered at the randomization block level. * indicates $p < 0.05$

Table B.3 Covariate Balance of Parent Characteristics, Sample of Boys

	(1)	(2)	(3)	(4)
	Combo	Math	F-Test (p-Value)	N
Parent Education				
Less Than High School	0.005	0.01	0.962	675
	(0.043)	(0.037)		
High School	0.048	0.011	0.504	675
	(0.042)	(0.040)		
Some College	-0.015	0.005	0.902	675
	(0.045)	(0.038)		
Associate's Degree	0.003	-0.004	0.945	675
	(0.026)	(0.022)		
Bachelor’s Degree	-0.02	-0.025	0.717	675
	(0.031)	(0.031)		
Master's Degree or Higher	0.012	0.038*	0.112	675
	(0.017)	(0.018)		
Missing Education Information	-0.033	-0.035	0.649	675
	(0.041)	(0.044)		
Parent Age	-1.238	-0.377	0.516	675
	(1.092)	(0.931)		
Parental Income	808.406	-2916.362	0.78	675
	(4908.648)	(4927.007)		
Hours Worked	-2.035	-3.505	0.627	675
	(3.659)	(3.646)		
Texting Language				
English	-0.058	-0.019	0.242	675
	(0.034)	(0.035)		
Spanish	0.041	0.02	0.4	675
	(0.030)	(0.027)		
Chinese	0.017	-0.001	0.5	675
	(0.016)	(0.021)		
Received Texts in 2015-2016	0.01	0.005	0.627	675
	(0.010)	(0.011)		
Received Texts in 2016-2017	-0.01	-0.001	0.89	675
	(0.028)	(0.035)		
Average Parent Baseline Reports of Attitudes and Activities				
Knows How to Support Literacy/Math/Behavior	-0.096	-0.012	0.785	387
	(0.163)	(0.200)		
Frequency of Literacy Related Activities	-0.234	-0.168	0.414	387
	(0.176)	(0.196)		
Frequency of Math Related Activities	-0.305	-0.194	0.272	389
	(0.188)	(0.177)		
Frequency of Behavior Related Activities	-0.062	-0.205	0.447	387
	(0.171)	(0.173)		

Notes: Parent baseline reports of attitudes and activities are domain averages of standardized items. Averages are standardized to have mean zero and standard deviation one. All models include randomization block fixed effects. Standard errors are clustered at the randomization block level. * indicates. $p < 0.05$

Table B.4: Covariate Balance of Student Characteristics, Sample of Girls

	(1)	(2)	(3)	(4)
	Combo	Math	F-Test (p-Value)	N
Age	-0.062	-0.007	0.123	661
	(0.033)	(0.033)		
Has 3 Year Old Sibling	-0.01	-0.006	0.635	661
	(0.010)	(0.014)		
Has Twin	0.017	0.011	0.521	661
	(0.016)	(0.014)		
Asian	-0.021	0.022	0.632	661
	(0.038)	(0.049)		
Black	0.035	-0.015	0.512	661
	(0.038)	(0.040)		
Hispanic	-0.04	-0.026	0.551	661
	(0.038)	(0.041)		
White	0.007	-0.01	0.8	661
	(0.031)	(0.028)		
Other	0.011	0.033	0.563	661
	(0.025)	(0.030)		
Missing Race/Ethnicity Information	0.012	-0.002	0.481	661
	(0.010)	(0.002)		
Fall DRDP				
Approaches To Learning	-0.002	0.036	0.915	644
	(0.101)	(0.112)		
Social and Emotional Development	0.029	-0.005	0.928	644
	(0.106)	(0.120)		
Language and Literacy Development	0.044	0.047	0.905	644
	(0.107)	(0.124)		
Cognitive Development	0.029	0.008	0.956	644
	(0.100)	(0.117)		
Physical Development and Health	0.116	0.02	0.483	644
	(0.105)	(0.116)		

Notes: Fall DRDP domain averages activities are standardized to have mean zero and standard deviation one. All models include randomization block fixed effects. Standard errors are clustered at the randomization block level. * indicates $p < 0.05$

Table B.5: Covariate Balance of Parent Characteristics, Sample of Girls

	(1)	(2)	(3)	(4)
	Combo	Math	F-Test (p-Value)	N
Parent Education				
Less Than High School	-0.015	-0.014	0.912	661
	(0.038)	(0.038)		
High School	-0.086	-0.07	0.155	661
	(0.044)	(0.052)		
Some College	0.006	0.03	0.793	661
	(0.040)	(0.048)		
Associate's Degree	0.06*	0.043	0.031	661
	(0.024)	(0.028)		
Bachelor’s Degree	0.002	0.013	0.882	661
	(0.028)	(0.026)		
Master's Degree or Higher	-0.031	-0.02	0.22	661
	(0.018)	(0.022)		
Missing Education Information	0.064	0.018	0.213	661
	(0.037)	(0.044)		
Parent Age	-0.974	-0.17	0.605	661
	(1.036)	(1.078)		
Parental Income	1661.345	5424.272	0.514	661
	(4095.036)	(4697.373)		
Hours Worked	-0.279	1.229	0.851	661
	(2.758)	(3.142)		
Texting Language				
English	0.034	0.028	0.529	661
	(0.032)	(0.030)		
Spanish	-0.014	0.02	0.695	661
	(0.021)	(0.027)		
Chinese	-0.02	-0.033	0.291	661
	(0.027)	(0.022)		
Received Texts in 2015-2016	0	0.005	0.934	661
	(0.017)	(0.018)		
Received Texts in 2016-2017	0.038	0.038	0.225	661
	(0.026)	(0.024)		
Average Parent Baseline Reports of Attitudes and Activities				
Knows How to Support Literacy/Math/Behavior	0.054	-0.015	0.807	661
	(0.109)	(0.135)		
Frequency of Literacy Related Activities	0.1	0.13	0.538	661
	(0.118)	(0.134)		
Frequency of Math Related Activities	0.179	0.066	0.356	661
	(0.127)	(0.146)		
Frequency of Behavior Related Activities	0.022	-0.044	0.858	661
	(0.124)	(0.134)		

Notes: Parent baseline reports of attitudes and activities are domain averages of standardized items. Averages are standardized to have mean zero and standard deviation one. All models include randomization block fixed effects. Standard errors are clustered at the randomization block level. * indicates $p < 0.05$

Table B.6: Attrition Balance, By Gender

	(1)	(2)
	Combination	Pure Math
Panel A: Boys		
Not Assessed	-0.033	-0.032
	(0.037)	(0.035)
Panel B: Girls		
Not Assessed	0.045	-0.015
	(0.040)	(0.039)

Notes: All models include randomization block fixed effects and a covariates detailed in Tables 3a and 3b. Standard errors are clustered at the randomization block level. N(Boys) = 945 and N(Girls = 897).

Table B.7: Alternative Specifications to Estimate the Effect of Combo and Math Program

	Mean Math Score				
	Combo		Math	p-Value (Math vs Combo)	N
Panel A: Main Specification (Covariates + FEs)					
All Students	0.000		-0.034	0.569	1336
	(0.058)		(0.056)		
Girls	0.156	+	0.015	0.160	661
	(0.083)		(0.099)		
Boys	-0.115		-0.025	0.324	675
	(0.105)		(0.094)		
Panel B: No Randomization Block FEs, Covariates					
All Students	0.022		-0.008	0.559	1336
	(0.051)		(0.051)		
Girls	0.161	*	0.038	0.087	661
	(0.072)		(0.072)		
Boys	-0.092		-0.032	0.425	675
	(0.076)		(0.075)		
Panel C: No Randomization Block FEs, No Covariates					
All Students	0.022		0.002	0.764	1336
	(0.067)		(0.068)		
Girls	0.200	*	0.083	0.191	661
	(0.090)		(0.092)		
Boys	-0.157		-0.077	0.415	675
	(0.099)		(0.098)		

*Notes: +, * correspond to p<0.1 and p<0.05. Standard errors in parenthesis. Mean math score is the average of the standardized subscores. The average is standardized to have mean zero and standard deviation one. In Panel B and C, standard errors are not clustered on the randomization block level.*