NBER WORKING PAPER SERIES

MORE THAN JUST A NUDGE: SUPPORTING KINDERGARTEN PARENTS WITH DIFFERENTIATED AND PERSONALIZED TEXT-MESSAGES

Christopher J. Doss
Erin M. Fahle
Susanna Loeb
Benjamin N. York

Working Paper 24450
http://www.nber.org/papers/w24450

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2018

We give special thanks to Carla Bryant, Meenoo Yashar, Pamela Geisler, and numerous other employees of the San Francisco Unified School District for the many ways in which they supported this study. The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B090016 to Stanford University as well as generous grants from the Silver Giving Foundation and the Evelyn and Walter Haas Jr. Foundation. The opinions expressed are those of the authors and do not necessarily represent views of the Institute or the U.S. Department of Education or the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w24450.ack

More than Just a Nudge: Supporting Kindergarten Parents with Differentiated and Personalized Text-Messages

Christopher J. Doss, Erin M. Fahle, Susanna Loeb, and Benjamin N. York

NBER Working Paper No. 24450
March 2018
JEL No. C93,D91,I21

ABSTRACT

Recent studies show that texting-based interventions can produce educational benefits in children across a range of ages. We study effects of a text-based program for parents of kindergarten children, distinguishing a general program from one adding differentiation and personalization based on each child's developmental level. Children in the differentiated and personalized program were 63 percent more likely to read at a higher level ($p<0.001$) compared to the general group; and their parents reported engaging more in literacy activities. Effects were driven by children further from average levels of baseline development indicating that the effects likely stemmed from text content.

Christopher J. Doss
Stanford University
cdoss@stanford.edu

Erin M. Fahle
Stanford University
520 Galvez Mall
Stanford, CA 94305
efahle@stanford.edu

Susanna Loeb
524 CERAS, 520 Galvez Mall
Stanford University
Stanford, CA 94305
and NBER
sloeb@stanford.edu

Benjamin N. York
ParentPowered Technologies
10 Mulberry Ct. #3
Belmont, CA 94002
ben.york@parentpowered.com

1. Introduction:

Educational interventions based on behavioral economics principles have shown promise for combatting some of the persistent disparities in education outcomes. Some of these interventions focus on helping participants hold their attention to tasks that need to be completed repeatedly over long periods of time (Bergman 2016). Others provide small bits of information regularly with easily operationalized tasks in order to overcome both information asymmetries and the cognitive load required for behavior change (York, Loeb, and Doss 2017). This information and support encourages parents and students to behave in ways that are more consistent with positive educational outcomes. Researchers have fielded successful interventions at all levels of education ranging from prekindergarten (York, Loeb, and Doss 2017), to K-12 (Kraft and Rogers 2015; Bergman 2016), to the transition to college (Hoxby and Turner 2013; Castleman and Page 2015). Such programs, due to their low cost and ease of implementation, provide researchers with opportunities not only to directly support students and parents but also to test the mechanisms underlying the effects of these programs.

This study aims to identify the importance of personalization and differentiation within a text-messaging program for parents of young children. Personalization conveys a combination of child-specific information and a potential increased sense of familiarity. It may provide parents with better information about their child and encourage a sense of connection that could lead to greater incentives for behavior change. Differentiation provides activities for parents that are targeted to their child's level of development and thus is potentially more effective for generating learning gains than a generic program. Differentiation may, in turn, encourage parents to engage more with the program if their children more successfully complete the developmentally appropriate activities. On the other hand, if program-inspired behavior change comes solely or

primarily from holding attention through regular reminders ("nudges"), we would not expect either differentiation or personalization to affect program effectiveness.

We field a randomized control trial to explicitly test the additional benefits of differentiating and personalizing information in a program for kindergarten parents modeled after the original READY4K! program for prekindergarten parents. The READY4K! program has been shown to increase the number of reported academic activities done at home and in turn to increase pre-literacy skills of children (York, Loeb, and Doss 2017). This study follows the first cohort of participants from the original experiment into their kindergarten year, recruits additional kindergarten families, and randomizes families to receive a small number of control text messages unrelated to literacy, general literacy texts, or literacy texts that are differentiated and personalized.

We employ a "light touch" differentiation and personalization that leverages extant data to adjust text messages. We personalize the texts by informing parents how well their child knew a particular skill based on the child's performance on formative assessments. We then differentiate the texts by aligning the activity more closely to the child's skill level. Through this experimental design we are able to test whether the differentiated and personalized information provision generates a greater parental response and greater academic gains in reading when compared to a general provision of information. Thus, we are able to identify the causal effect of differentiation and personalization separate from the effect of information provision alone.

We find that differentiation and personalization increases parental take up of the program as measured by parental survey responses and increases the reading ability of students as measured by district assessments. Specifically, differentiation and personalization caused students to be 63 percent more likely to move up a reading level than their peers in the general program ($p<0.001$), with the academic effects particularly pronounced for students in the bottom and top quartiles of

the baseline skill distribution. The differentiated and personalized texts also positively affected parents' reports of the ease of building reading skills by 32 percent of a standard deviation ($p<0.05$) when compared to the general texting program, while increasing parental engagement in literacy activities with their child by 26 percent of a standard deviation ($p<0.05$) when compared to the control group. There is some suggestive evidence that the differentiation caused parents to use the texts more, indicating that a closer match of the text to the children's skill level led parents to engage in the activities to a greater extent. The greater amount of information in the texts, however, may have caused parents to visit their children's school less often.

2. Background

Recent experiments in education have demonstrated that parent texting interventions based on behavioral economics principals are effective in improving students' educational outcomes. The precursor to this study, the READY4K! experiment conducted in the prekindergarten context, applied several of these principals. Families in the treatment group received three literacy texts per week for eight months. The program provided a parenting curriculum that was designed to remedy information asymmetries and limited attention, breakdown the cognitively complex task of engaging in academic tasks with small, easy-to-achieve activities, provide encouragement, and reframe distal rewards to be more proximal. Though the evidence of information asymmetries is mixed (Avery and Kane 2004; Grodsky and Jones 2007; Hastings and Weinstein 2008; Valant and Loeb 2014), suboptimal behavior due to limited attention (Karlan et al. 2016), the cognitive complexity of tasks (Mullainathan and Thaler 2000), and time inconsistent preferences (DellaVigna 2009) is well established. Addressing these behavioral barriers with the READY4K! program yielded substantial literacy benefits. The program, implemented in the San Francisco Unified School District (SFUSD), increased take up of home literacy activities and parental

involvement in schools by approximately 20 percent of a standard deviation and increased some pre-literacy test scores by approximately 10 percent of a standard deviation (York, Loeb and Doss 2017).

Addressing behavioral barriers through texting has been applied to a range of levels of education. These interventions often include child-specific information to address information asymmetries and limited attention. On the K-12 level, Bergman (2016) used email, text messages, and phone calls to inform parents of their child's missing assignments. The information given was student-specific and detailed, often containing specific class assignments and page numbers, and clearly personalized for specific parents and students. The intervention led to a 21 percent of a standard deviation increase in student GPA, a 25 percent increase in assignment completion, and a 28 percent decrease in classes missed. Kraft and Rogers (2015) used the same three mediums to establish weekly teacher-parent communication in the summer school context. In one treatment arm teachers conveyed positive messages regarding their child's behavior and academic performance. In another treatment arm, teachers highlighted areas where the child could improve. The authors found that this intervention increased the probability of passing the summer school class by 6.5 percentage points – a 41 percent reduction in failing the class. The results were driven mostly by the child-specific suggestions parents received on where the child could improve. Receiving positive information regarding child-specific successes produced positive, though imprecise, point estimates.

At the post-secondary level. Castleman and Page (2015) fielded an intervention to help ease the transition to college for new high school graduates. A text messaging arm of the treatment sent differentiated and personalized reminders during the summer regarding deadlines for filling out the required paperwork to matriculate into college. The information in the messages was

specific to the requirements of the college in which the student was accepted and planned to matriculate. Students received reminders to access important paperwork, register for orientation, register for placement tests, complete housing forms, and complete health insurance forms. A second treatment arm used in-person peer mentors that reached out to students directly to offer help in completing the required tasks. Both treatments increased college enrollment among students who had less access to college counseling during the academic year.

Though this line of literature is compelling, it is unclear which elements of the programs are driving the results. The current study seeks to test whether these texting programs are effective because they address limited attention through reminders (a "nudge") or through the other behavioral barriers that require a greater interaction with the content, which is often personalized and differentiated to program participants. In this vein, we test whether differentiating and personalizing the READY4K! intervention increases (or decreases) program effectiveness. We personalize the text messages by providing information to parents about their child's skill level, as measured by formative assessments already administered by the district. We differentiate the text messages by providing parents a literacy activity tailored roughly to their child's skill level. A significant, differential effect of the personalized and differentiated version of the text messages will provide evidence that parents are actively engaged with the program content. The program can reduce the cognitive load inherent in parenting, provide novel information to parents, and address time-inconsistent preferences only if parents absorb the content of the messages. If they do not, then the positive effects of the texting programs are likely driven simply by nudges that hold attention.

Interaction with the content opens the door to the possibility that personalization and differentiation uniquely affects parent behavior. There are many different channels through which

this effect can occur. The mere knowledge that the texts are tailored to a child may induce parents to engage with the texting program more regularly. To our knowledge, no study has tested to see whether personalization of interventions engenders more trust and fidelity to treatment from participants. However, behavioral economics has produced a robust line of literature that shows that how information is presented to people affects subsequent behavior. For example, the social norms literature shows that presenting someone with information on their peers' behavior can lead to lower energy use (Allcott 2011) and increased savings (Kast et al. 2012), charitable giving (Frey and Meier 2004), and voter turnout (Gerber and Rogers 2009). The text messages in this experiment do not provide information about the behavior of the parents' peers, but do provide information about the parents' children. The close relationship between parents and children may amply the effects seen in the larger social norms literature.

The information conveyed about parents' children's performance on formative assessments may also update inaccurate beliefs regarding their children's ability. For example, parents may not realize their children are weak or strong on certain literacy skills and therefore fail to invest in their children's development efficiently. In Malawi, Dizon-Ross (2017) illustrates that parents inefficiently invest in their child's academic success due to inaccuracies in their perceptions of their child's ability. The gap between perceived and actual ability is as large as one standard deviation. After receiving information specific to their children, parents began to invest in their children more efficiently. Their willingness to pay for remedial materials decreased as a function of their children's revealed performance and they more accurately picked textbooks that matched to their children's ability. One year later the higher forming children were more likely to transfer to better schools and were less likely to drop out, though lower performing siblings were hurt academically in the process.

Finally, the closer alignment of the difficulty of the task to the child's skill may lead to a greater probability of success in carrying out the activities. This success may in turn produce a recursive feedback mechanism that encourages parents to continue with the program. Lower performing children who previously received tips that were too advanced may have failed at the activities, causing parents to disengage from the program. Advanced children who, in the counterfactual, receive activities that are too easy, may have gained little from the experience, causing parents to disengage from the program. Again, no study has specifically probed this mechanism.

Apart from inducing behavioral changes in parents, aligning the difficulty of the activity to the ability of the child may also produce differential academic gains. Education theorists posit that students advance in knowledge when taught concepts that are slightly beyond, but still close to, the student's ability, a concept called the "Zone of Proximal Development" (ZPD) (Vygotsky 1978a, 1978b). Traditionally, practitioners have used formative assessments to provide information on where a child's ZPD lies and have grouped students by ability to tailor instruction and activities to students whose ZPD lie in approximately the same place. Many studies have shown that the use of formative assessments and data can improve the educational outcomes of children. In a meta-analysis of studies conducted between 1988 and 1998, Black and William (1998a, 1998b) find that the use of formative assessments can increase student performance by 40 to 70 percent of a standard deviation, with effects prominent for low-performing children. In kindergarten, benefits of using formative assessments have been seen in reading, math, and science outcomes (Bergen and Sladeczek 1991). Ability grouping in the classroom context has also generally led to positive academic results for children by as much as a quarter of a standard deviation (Kulik and Kulik 1992, 1982; Robinson 2009). The effects, however, are not uniformly

positive, with some evidence that lower performing children can be hurt through ability grouping (Lou et al. 1996).

With the rise of artificial intelligence, technology has been used to more efficiently identify a child's ZPD so that instruction can be tailored to the child. As the software gathers information on the child's ability it tailors the program and activities to be more aligned to the child's skill. As the child's skill changes and grows, the software adapts the educational outputs accordingly. This process also provides teachers information that can be used to differentiate instruction. The evidence on these types of software is mixed. Van Klaveren and colleagues (2017) compare test score outcomes of Dutch children randomly assigned to either a static or adaptive practice technology and find no overall benefits to the adaptive technology and negative effects of about 8 percent of a standard deviation on higher performing children. Pane and colleges (2014) analyze the efficacy of the Cognitive Tutor Algebra I program that provides a curriculum designed around an adaptive technology software. They found mixed results when comparing children in schools randomly assigned to the curriculum to children in traditional Algebra I classrooms. Positive results appear for high schoolers in the second year of the experiment. These two studies were included in a recent review of educational technologies by Escueta et al. (2017). Across 29 studies, she finds a similar mix of results, with some interventions providing large benefits to children, and others producing null results.

To the extent that the differentiation and personalization of text messages produce differences in outcomes, we will be unable to disentangle the effect of differentiation from personalization. We will also be unable specifically test for the behavioral mechanisms through which differentiation and personalization can act. However, indirect evidence on this latter point can be culled by analyzing responses on parent survey items. We also assess whether the

intervention was more successful for students of average baseline ability or for students who started at the tails of the baseline ability where differentiation was greatest, potentially distinguishing the effects of personalization engendering a feeling of familiarity from the other potential mechanisms.

3. The Intervention

This study is an extension of the READY4K! intervention run in conjunction with the San Francisco Unified School District (SFUSD) starting in the 2013-2014 school year. In the original program, treatment families received three texts per week. The "FACT" text was sent on Mondays and informed families of the skill of the week and the importance of that skill for the academic growth of their child. On Wednesdays families received a "TIP" message that suggested an home literacy activity based on that skill. These literacy activities were meant to fit as seamlessly as possible into the parents' day and to capitalize on items and materials found in their home and neighborhood. These "TIP" texts aimed to provide an easy choice to parents and thus reduce the cognitive load inherent in parenting that stems from making multiple and ambiguous choices. Finally, on Fridays, families received a "GROWTH" text that contained a more advanced activity that was meant to extend the learning opportunity presented earlier in the week as well as encouragement aimed to provide some immediate gratification. Control families received one text every two weeks that contained general district information and did not promote parent-child interactions. The eight-month long program touched on a variety of pre-literacy skills such as letter recognition, letter sounds, rhyming, and early literacy behaviors. Participants could choose to receive the texts in English, Spanish, or Chinese (York, Loeb, and Doss 2017). For this study, we built on the original READY4K! format of FACT/TIP/GROWTH but created new texts to match

the skills covered in kindergarten. We created both a generic version and a differentiated and personalized version targeted to students' developmental level.

To field the study, we followed the first cohort from the original experiment into their kindergarten year and recruited more of their kindergarten peers. The original participants in the control condition remained in the control condition in the second year. The original participants in the treatment condition were re-randomized to either continue receiving general literacy texts or to receive differentiated and personalized literacy texts. Newly recruited participants were randomized to receive either general texts, differentiated and personalized texts, or control texts. To keep the proportion of families treated the same in each cohort, half the new participants were randomized to receive control texts, and half were randomized into the two treatment arms.

To recruit new participants in the study, we worked with parent liaisons in each elementary school. In August 2014, we provided a brief training to liaisons to explain the study, its purpose, and provide materials with which to recruit families. Through their regular course of business, liaisons recruited families to participate in the study.[4] Families that consented to participate completed a baseline survey to elucidate their home literacy habits and the skill level of their children on a variety of literacy skills. We used some of the same questions from the baseline survey in the original year so that we could pool answers between cohorts and use the responses as covariates in an effort to increase the precision of our estimates. As an incentive, liaisons were paid $10 for every family they recruited into the study. Participants in both the treatment and

[4] The job of a parent liaison to facilitate communication between families and the school. One responsibility of parent liaisons is to coordinate school services to students and parents and refer families to school resources. It is during this process that parent liaisons recruited parents for the study. If a parent liaison talked to a parent in the context of coordinating or referring services, they informed the parent of the program. If the parent consented they then completed the consent form and baseline survey.

control conditions were paid $10 a month as long as they remained in the program, with the aim of covering texting costs for parents without unlimited texting plans.

We began texting at the end of October 2014 and continued for ten months. We used fall first grade literacy assessments as the primary outcome of interest. Details regarding the three randomized conditions are as follows:

(1) *Differentiated and Personalized Text Treatment:* Treatment followed the same general design as the first year of the experiment. Families received three texts a week: a "FACT" text on Mondays, a "TIP" text on Wednesdays, and a "GROWTH" text on Fridays. Only the TIP and GROWTH texts were differentiated and personalized using child level formative assessment data on skills that corresponded to the week's topic. The literacy texts reviewed skills from prekindergarten such as letters, letter sounds, and rhyming. Then they eased parents into asking their child to read and helped parents teach their children to read with greater accuracy and comprehension. Figure 1 presents the differentiated and personalized versions of the texts (see Figure A1 in the online appendix for additional examples).

We insert two pieces of information in the TIP texts. First, we personalize the texts by giving parents an indication as to where their child falls in the distribution of skills. As seen on Figure 1 we indicate that the child is "beginning" to learn the skill, "growing" in their knowledge of the skill, has a "solid" understanding, or has a "strong" knowledge of the skill. We positively framed each text so that parents of children on the lower end of the distribution would not become frustrated. This framing is akin to an "injuctive norm" in the behavioral economics literature. Additionally, the texts are differentiated such that parents receive one of four different activities based on their child's prior academic information. At first we used parental responses from the baseline surveys, and once available, we switched to data from the fall, winter, and spring

administrations of the Fountas and Pinnell Benchmark Assessment System (BAS) administered by the child's kindergarten teacher. We identified the relevant skill for each week as measured by the BAS and divided the skill's scale into four equal intervals. Students scoring in each interval received different TIP and GROWTH texts. Those on the lower end of the distribution received easier versions of the TIP and GROWTH, while those at the upper end of the distribution received more advanced versions. A child was not necessarily in the same category each week because a child may be weaker on one skill, but stronger in another.[5] The information in the TIP text often would not fit into one text. In those cases, families received two texts on Wednesday, one right after the other. As a result, families in this condition received one extra text message per week, though the timing and spacing of texts was very similar across treatment groups.[6]

(2) *General Text Treatment:* The families randomized into the general text treatment also received FACT, TIP, and GROWTH texts each week. The FACT texts were identical to those received in the differentiated and personalized text condition. The TIP and GROWTH texts, however, did not include the strength of their children on the particular skill, and every family in this condition received the same activity. The activity was most often similar to, if not identical to,

[5] While we explicitly informed parents that the tip was based on their child's formative assessment performance, we did not explicitly indicate to parents that "beginning," "growing," "solid," and "advanced" were terms that indicated a child's performance along a continuum of skill levels. Parents could have deduced the implied meaning of these words if their child fell in different groups over time, across skills. This was the case for almost all parents. Only 1 child was consistently in a group throughout the experiment.

[6] There may be an effect of receiving one extra text per week, in addition to the personalization and differentiation. In other work, we are explicitly testing the effect of receiving more TIPS during the week. In one treatment arm recipients received a FACT/TIP/GROWTH program akin to the general program in this study. In another treatment arm recipients received a FACT/TIP/TIP/TIP/GROWTH program. Preliminary results show no differences in outcomes but survey responses were slightly less positive for the group that received five texts. These results indicate that the extra text received in this outcome likely did not affect academic outcomes and may have slightly attenuated survey results (Cortes et al 2017). Results available on request.

the activity given to families in one of the middle two groups in the differentiated and personalized text treatment. This treatment condition is directly analogous to the original texting experiment. Figures 1 and A1 (in the online appendix) give examples of the general texts.

(3) *Control Text Condition:* Families in the control condition received one text, every other week, with information about the school district. The two examples presented in Figure 1 provide information on emergency preparedness and on how the food in SFUSD is prepared.

For all conditions, parents could choose to receive the texts in English, Spanish, or Chinese.

4. Data and Empirical Strategy

4a. Data and Sample

The initial sample included 504 children and families from the original experiment and 290 newly recruited children and families. These 794 students were randomized into one of the three conditions and received texts from October 2014 through August 2015. We collected three primary sources of outcome data on these children. In May 2015, we surveyed the kindergarten teachers of all the children in the study. We asked questions regarding how well the teacher knew the parents of the children, how often parents talked to the teacher, how often parents asked questions regarding specific academic skills, and how well the child performed on specific academic skills. Teachers were not informed of the treatment status of individual children so as not to bias the results. Each teacher received $50 for completing the survey.

In September 2015, after texting was complete, we sent parents enrolled in the program a post-survey. We asked questions regarding their attitudes towards building literacy skills in their children, how often they engaged in specific learning activities with their children, how often they interacted with their children's teacher, and how they viewed the texts they received. We also compensated parents $50 for completing the survey.

Finally, we use the fall first grade administration of the Fountas and Pinnell Benmark Assessment System (BAS) as a measure of children's academic skills. The BAS is a formative assessment tool that has been shown to be a valid assessment of literacy development in children (Fountas and Pinnell 2012). Teachers first assess the ability of children to recognize upper-case and lower-case letters, letter sounds, initial word sounds, 25 high frequency words, rhyme, blend sounds into words, and demonstrate early literacy behaviors. After mastering six of the eight foundational skills children are asked to read books of increasing difficulty. The teacher begins with the easiest books, level A. After the child reads with sufficient accuracy and comprehension, they move on to harder books (levels B-Z). A teacher stops after reaching a book that the child cannot read with sufficient accuracy and comprehension. In kindergarten, most children are still mastering foundational skills, while in first grade the vast majority of children are reading books of varying difficulty. The texts are therefore primarily differentiated based on a child's performance on the eight foundational skills listed above. The outcome of interest is whether children are reading more complex books in the fall of first grade and whether they reach development benchmarks set forth by Fountas and Pinnell.

Of the 794 participating families, teachers provided information on 442 (56% response rate) students, 519 families responded to the survey (65% response rate), and 641 students completed the fall first grade BAS (81% assessment rate). The 153 students who do not have assessment data left the district. This level of mobility in the early grades is not uncommon. Only 28 children that we recruited in the beginning of the year left during the year or in transition to first grade. The remaining 125 children are from the original cohort recruited during SFUSD's prekindergarten enrollment process. These students left the district between enrolling for prekindergarten and transitioning to first grade. To obtain the final analytical sample we restrict

the sample to those parents who answered enough baseline survey questions to construct three measures of their pre-treatment characteristics. The baseline survey was designed to measure three constructs: baseline child skills, baseline frequency of literacy activities in the home, and background characteristics of the parent. To get a measure of each construct, we estimated a graded response model separately on each subsection of the survey. Graded response models (GRMs) are used frequently in survey analysis with Likert-type items, and provide an estimate for all respondents of where they fall along the construct of interest, termed their "ability" estimate. We selected GRMs over factor analysis due to their ability to produce estimates in the presence missing data, avoiding imputation (Samejima 1997). We use these ability estimates, rather than the individual questions, as control variables in our analyses.

In the end, we have three analytic samples. The final parent survey sample consists of 475 families, the teacher survey sample consists of 409 children, and the BAS sample consists of 578 children. We check to ensure that attrition and pre-treatment covariates remain balanced in all samples. Finally, we merge this data to district administrative data on student background characteristics such gender, ethnicity, and date of birth.

Table 1 presents the descriptive statistics on each analytic sample. Though there are slight differences among samples, the demographics are largely similar. Panel A presents the characteristics of children in the sample. Looking at the academic outcome sample, 51% of the children are male. The two largest ethnicities are Hispanic (35%) and Asian (33%), with fewer white children (7%) and children from other ethnicities (19%). The average age in the sample is 5.4 years old. At baseline parents rated their children 3 out of 4 in letter knowledge and a 3 out of 5 in letter sounds and rhyming, on average. In comparison, the broader SFUSD kindergarten cohort has more white students (14 percent) and fewer Hispanic and Asian students (27 percent and 23

percent, respectively). Both samples, however, have approximately the same proportion of males, students from other ethnicities, and students of approximately the same age.

Panel B presents descriptive statistics on the parents. Most have less than a bachelor's degree (65%) and are on average 34 years old. Over half (53%) chose to receive the text messages in English, with fewer choosing to receive them in Spanish (26%) and Chinese (22%). A little less than half the sample (42%) is new to the program this year. On average parents rated themselves between 2.8 and 3 out of 4 when asked how frequently they engage in literacy activities with the child. The texting program primarily served non-white and lower-income families.

4b. Empirical Strategy

We use the following model to estimate the effect of the texting program on student and family outcomes:

$$(1)\ Y_{is} = \beta_0 + \beta_1 GeneralText_{is} + \beta_2 PersonalizedText_{is} + \boldsymbol{X}_{is}\beta_3 + \alpha_s + \varepsilon_{is}$$

In Equation 1 we regress an outcome Y_{is} for student, i, in school, s, on $GeneralText_{is}$, an indicator for receiving the general literacy texts, $PersonalizedText_{is}$, an indicator for receiving differentiated and personalized texts, $\boldsymbol{X_{is}}$, a vector of baseline characteristics, and α_s, a school fixed effect. ε_{is} is a stochastic error term. $\boldsymbol{X_{is}}$ contains an indicator for receiving the texts in English, Spanish, or Chinese, the child's gender, ethnicity, age in years, and factors of baseline survey questions on literacy skills and rates of home literacy activities. Randomization occurred within school site, and the school fixed effect, α_s, is the school site where randomization took place. For children in their second year of the experiment this is their prekindergarten school site and for children in their first year of the experiment this is their kindergarten school site. First or second year status does not vary within randomization school sites. We therefore do not include an indicator for being new to the experiment in $\boldsymbol{X_{is}}$. The coefficients of interest are β_1 and β_2, which provide estimates of the

effect of receiving general and personalized and differentiated texts, respectively, on the outcome of interest. The omitted group in this case is the control group. We cluster all standard errors at the randomization site level.

In supplementary analyses, we replace the indicator for receiving general texts with an indicator for receiving *any* literacy text, $AnyLiteracyText_{is}$, which is equal to one for students in either the general text treatment or the differentiated and personalized text treatment. All other elements of the equation remain the same. In this specification, β_2 provides an estimate of the effect of the personalization and differentiation, relative to the effect of receiving general text messages. In an effort to be parsimonious, we do not present the results of this model. Its main advantage is that it indicates whether the difference in effects between the general text messages and the personalized and differentiated text messages are statistically significant. We reference the significance in the body of the paper when relevant.[7]

The outcomes, Y_{is}, are the individual teacher and parent survey questions and the reading level of the child as measured by the BAS. To reduce the number of outcomes from the surveys we use exploratory factor analysis to determine which questions measure the same underlying construct. The questions in the parent survey load onto four separate factors: (1) a parent belief factor regarding the ease of building literacy skills including the support they feel in building those skills, (2) a literacy activity factor capturing the frequency with which the parents engage in literacy activities with their child, (3) a teacher factor regarding the frequency with which parents interact with their child's teacher, and (4) a text factor regarding parental attitudes to the texting program. In creating the final factors, we used principal components analysis and rotated the loading matrix to create orthogonal factors. For the teacher survey, we took the analogous

[7] Results are available upon request

questions from the parent survey and created a teacher version of that factor, so that the two are directly comparable. Table A1 in the online appendix presents each question contained in each of these factors and the weighting of the elements variables for each of these factors.

4c. Randomization Checks

The covariates are largely balanced between treatment and control for each analytic sample. Table A2 in the online appendix presents these results for 14 covariates tested in each of the three samples, for a total of 42 tests. No variable in the parent survey sample was significantly unbalanced. In the teacher survey sample, one variable was unbalanced at the 10 percent level (male) and one variable was unbalanced at the 5 percent level (white). In the academic sample, one variable was unbalanced at the 1 percent level (white). The rate of imbalance is about what one would expect to occur by chance in the parent survey and academic samples, but is a little higher in the teacher survey sample. All our main specifications include covariates and we present all results with and without covariates. For all outcomes, addition of the covariates does little to change the point estimates, and ultimately does not change our inferences, providing an indication that imbalance is not a concern in this study.

We also test whether students differentially left the analytic samples. Table 2 shows that, overall, we do not find evidence that students differentially attrited from the parent survey sample or the academic sample. Attrition is marginally significant for the personalization and differentiation treatment arm in the teacher sample (-7.5 percentage points). We further check to see whether measurably different children left the sample. Appendix Table A3 in the online appendix shows no imbalance in the academic and parent survey samples, but older students are marginally less likely to attrit from the general texting group in teacher survey sample. Because the point estimates on the overall attrition is larger (and marginally significant) and the attrition by

age is marginally imbalanced in the teacher sample, we implement Lee (2009) style bounds on the teacher survey sample as a robustness check.

5. Main Results

Tables 3 through 5 present the main results of the intervention and show that the differentiated and personalized texts had positive effects. Table 3 presents the results on the fall first grade Fountas and Pinnell BAS. Panel A shows the effects of the program on the reading level of children, with level A being the easiest book and level Z being hardest book. A small minority of children (8%) were not yet reading. We analyze the results in three ways. First,
we capitalize on the ordinal nature of the reading scale and use ordinal logit models. To aid the interpretability of the results we create a standardized, linear scale from the reading levels and present results as effect sizes. We also present the results of linear probability models that show the effect on the probability of reading at level A, C, E, or G and above. Level A indicates that the child is first able to read, and levels C, E, and G represent the 25^{th}, 50^{th}, and 75^{th} percentiles of the reading distribution. Panel B presents the effects of the program on the probability of meeting district benchmarks. These benchmarks track the Fountas and Pinnell recommended benchmarks. Levels C, D, and E are the cutoffs for approaches, meets, and exceeds expectations. The academic results indicate that differentiated and personalized text messages had a significant effect on the reading ability of children while the general texts did not. Children whose parents received the differentiated and personalized texts messages were 63 percent more likely than the control group to move up a reading level ($p<0.001$). This estimate translates to an 18 percent of standard deviation increase in reading level ($p<0.05$). These children were also 8.84 percentage points ($p<0.05$) more likely to read at level E or above, were 12.05 percentage points ($p<0.001$) more

likely to exceed expectations, and 9.00 percentage points ($p<0.01$) more likely to meet or exceed expectations.

Tables 4 and 5 present the survey results which give clues to the mechanisms underlying the academic results. We present the results of the factors of survey questions in Panel A, as well as the results of individual questions in Panels B and C. Table A1 in the online appendix gives all questions that compose each factor. Table 4 shows that overall the texting program had limited effects on parent beliefs towards activities and building academic skills in their children. The program had the greatest effects on parent ratings of how easy it is to build literacy skills in their children. The general texting treatment caused a marginally significant, 27 percent of a standard deviation *reduction* in parent ratings on the extent to which they feel building literacy skills is easy. The differentiated and personalized intervention significantly mitigated these negative effects. Parents in the personalized treatment group responded 32 percent of a standard deviation ($p<0.05$) more positively than parents in the general treatment group. These results are consistent with the notion that knowledge of a child's skill level, with an appropriately differentiated activity, can positively affect parent beliefs. If general text messages were unaligned a child's skill level and too hard for parents, they could cause parents to believe that building literacy skills is difficult.

The program had a stronger effect on the frequency with which parents engaged in home literacy activities. Panel C of Table 4 shows that the greatest effects are in reading words with children, taking books when leaving the house, reviewing parts of a book, reviewing the direction of reading, correcting mistakes while reading, and practicing rhyming, with effect sizes ranging from 24 to 37 percent of a standard deviation. Differentiation and personalization drove some of these results. When combining all activities questions into a factor, Panel A of Table 4 shows that that general texts had a positive, but insignificant point estimate of 14 percent of a

standard deviation and that personalized texts had a significant 26 percent of a standard deviation ($p<0.05$) effect on home activities compared to the control group. We do not have the power to separate a differentiation and personalization effect from a base texting effect for this composite variable.

Table 5 presents the results of the intervention on parent involvement at school both from the parent perspective (Panel B) and the teacher perspective (Panel C). From the parent perspective, the largest effect is seen on how well they know their child's teacher and how often they inquire about how their child is getting along with other children. Columns 3 and 4 indicate that, compared to the control group, the general texts increased each of those two dimensions by 26 percent of a standard deviation ($p<0.05$) but the personalized texts did not have a significant effect. The differentiated and personalization aspect of the treatment marginally significantly decreased the positive effect on knowing the teacher generated by the general texts by 21 percent of a standard deviation when compared to the general text message group. The remainder of the estimates shows that the general texts increased specific questions parents asked teachers by about 10 to 20 percent of a standard deviation. The frequency with which parents talk to teachers about their child's interests and literacy skills reach marginal significance. The point estimates on the personalized and differentiated treatment in Table 5 are generally of equal magnitude or smaller compared to the general texts. Only one question, the frequency with which parents ask how well their child is doing in school reaches marginal significance with a point estimate of 17 percent of a standard deviation. When combining these measures into one factor in Panel A the pattern remains. General texts have a larger effect on teacher interactions of 23 percent of standard deviation ($p<0.10$), while personalized texts had a smaller, insignificant effect of 11 percent of a standard deviation.

The results are fairly consistent when analyzing the same questions from the teacher perspective. Column 3 of Panel C in Table 5 shows that the general text treatment had positive effects on parents talking to their child's teacher about their child's interest literacy skills, and home activities with effect sizes of 23 percent of a standard deviation ($p<0.10$), 25 percent of a standard deviation ($p<0.10$) and 28 percent of a standard deviation ($p<0.05$) respectively. The remainder of the point estimates are generally positive, but not significant. Column 4 however shows that the point estimates on the personalized texts are negative, with the effect on teachers knowing the parents reaching a marginally significant -26 percent of standard deviation. The differentiation and personalization texts lead to an estimated 32 to 43 percent of a standard deviation less positive effect for many questions when compared to the general text messages. This negative effect is also seen in the composite of the teacher reports in Panel A. The point estimates indicate that the personalization of the texts may have induced parents to talk to teachers less when compared with the general text messages and mute any gains in teacher interactions generated by the general text messages. These results, as well as the results from the parent questions, are plausible if the greater amount of information regarding the child skill level, in combination with greater success in implementing the differentiated literacy activity, produced less of an incentive to talk to the teacher regarding how their child is progressing in school.[8]

[8] The more positive reports of teacher interactions from parents of the personalized and differentiated group when compared to the teacher reports on the same questions may indicate that social desirability bias is greater for personalized and differentiated text recipients. While this is a possibility, this phenomenon would further support the assertion that parents are interacting with the content of the texts. In addition, the larger effects of the personalized and differentiated texts on academic outcomes indicate that larger survey point estimates are not completely driven by social desirability bias and parents are, in fact, engaging in activities to a greater extent compared to the general texting group.

Despite the potential decreases in interaction with the school, the greater extent to which parents engaged in activities in the personalization and differentiation group likely led to the academic advantages seen in Table 3. Unclear, however, is whether the knowledge that the texts were personalized engendered a greater fidelity to the program, whether updating parent beliefs about the child's skill caused a more efficient allocation of resources to the child, or whether a closer match between the difficulty of the texts and the skill level of the child led to greater success in carrying out the tips and established a positive, recursive feedback mechanism that encouraged parents to continue to engage in the texts. Parents in both groups may have interacted with the activities in similar ways and the closer alignment of the differentiated text to the child itself may have been the sole cause of the increased academic skills.

Though we cannot definitively identify which mechanisms are at play, responses to questions that elicited parental attitudes towards the text messages can provide some clues. Table 6 provides these results. We asked parents the extent to which they thought the texts were made for them and their children. Interestingly, there is little difference in response between the two groups that received the treatment texts. Both sets of parents reported the texts were made for their children to a much higher degree than the control group, with effect sizes around 40 percent of a standard deviation ($p<0.001$). Personalized and differentiated texts message did not elicit a greater response on this dimension. These results indicate that parents in the personalization and differentiation group did not see the text messages as more tailored to their child. Without this realization, personalization likely did not induce parents to adhere to the program more faithfully,

nor were they likely to more efficiently allocate resources to their children after receiving information on their child's skill level. [9]

We see greater differences in point estimates when we asked parents to what extent they used the texts, thought the texts were helpful, and would recommend the texts. In the case of using and recommending texts, the personalized and differentiated texts had a 29 and 23 percent of a standard deviation effect, respectively ($p<0.05$), while the general text messages had positive, but insignificant effect. In the case of the texts being helpful, both groups reported a significant effect compared to the control group. For each of those three questions the personalized and differentiated group effects were about 15 percent of a standard deviation higher than the general texting group, though we do not have the power to determine if this difference is significant. Though we must be extremely cautious in interpreting insignificant differences, these results do not eliminate the possibility that a greater success the personalized and differentiated text messages caused parents to engage with the program more faithfully.

6. Heterogeneity of Results

Prior research on social information experiments indicates that the effects of such interventions can vary significantly by baseline characteristics. Allcott (2011) demonstrates that providing families with information on their neighbor's energy usage will, on average, decrease their own energy use. Perhaps predictably, the effects are concentrated on the highest pre-intervention energy users, with no effects seen on the lowest pre-intervention energy users. Gerber and Rogers (2009) illustrate that presenting voters with a script that frames an upcoming election as a "high turnout" election will, on average, induce people to vote more compared to a script that

[9] There is no effect of the program on the rate with which the texts were read, indicating that all parents received the texts and even the control group read the texts.

frames the election as a "low turnout" election. They present evidence that the intervention was more effective for participants who voted less frequently in prior elections. Beshears et al. (2015) present a more nuanced result and show that the same intervention can have opposite effects depending on where participants fell in the baseline distribution. Their intervention provided social behavior regarding 401(k) savings and found that the intervention encouraged those who were previously contributing at high rates to save more, but discouraged those who were not previously saving much from contributing to their plans.

We analyze heterogeneity by the baseline skill distribution.[10,11] Specifically we estimate the effects of the intervention separately on students who fall in the middle two quartiles of the baseline skill distribution and on students who fall in the tails of the distribution. To do this analysis we must restrict the sample to those families who are new to the program. York, Loeb and Doss (2017) showed that texting in the first year positively affected pre-literacy skills. Because we are retaining the control group in this analysis, there will be a positive correlation between fall kindergarten test scores and texting treatment status for those families in the second year of the program.

[10] We also have analyzed the results by texting language. Splitting the sample into three languages greatly reduces the power to detect effects. Generally, the children of the parents receiving texts in Chinese saw the greatest academic gains. Results available upon request.

[11] We find little heterogeneity in academic outcomes by year of participation in the program. Both sets of families benefited to about the same extent, with differentiation and personalization driving the results. Effects are slightly higher for first year families, but are not significantly different than effects on returning families. If the first-year participants saw larger effects, this could be a result of texting fatigue in second-year participants. If this program was rolled out to a new, comparable, population of recipients we might expect to see slightly larger effects. However, because the inferences remain stable between years, we would expect personalization and differentiation to provides benefits regardless of whether participants received texts in the previous year. Academic results by length of time in the program are shown in Table A4 in the online appendix.

Previous results provide some evidence that the differentiation of the texts is driving the results. If this were the case, one would expect the effects to be concentrated on the tails of the distribution where differentiation is greatest. General texts most often corresponded to texts sent to either the second or third quartile. It is also possible that parents at the tails of the distribution respond to the personalization more strongly. In this case, the effects are unclear ex ante. On one hand the personalization may have greater effects on the tails of the distribution if parents are particularly motivated by signals that their child is doing relatively well or poorly. Similarly, the intervention could have smaller, or negative, effects at tails of the distribution if parents on the low end of the distribution are discouraged by the knowledge that their child is doing relatively poorly, or if parents at the top of the distribution feel less compelled to engage in the activities after learning their child is already advanced.

Table 7 shows the results of the heterogeneity analysis on the academic outcomes.[12,13] Panel A presents effects on the middle two quartiles and Panel B presents the effects on the first and fourth quartiles. Column 3 of Panel A in Table 7 shows that there is a marginally statistically significant effect of the general texts on the probability of meeting or exceeding expectations of 17.13 percentage points ($p<0.10$). Differentiation and personalization, however, produce no differential effect with a quantitatively similar coefficient in Column 4. Meanwhile neither treatment texting intervention had a detectable effect on the probability of approaching expectations or exceeding expectations. The ordinal logit model is imprecisely estimated. The results are quite different in Panel B, which presents results for families whose children are in the

[12] Sample sizes are too small to draw conclusions from the parent and teacher survey samples.
[13] Table A5 in the online appendix presents descriptive statistics on the sample by quartile. Families of children in the lower quartiles were more likely to be Hispanic and less likely to be Asian and white. Parents rated their children lower on baseline skills and reported engaging in literacy activities less frequently in the home. They were also less educated.

first and fourth quartiles of baseline academic skills. Column 3 and 4 indicate that the general texts had no effect on the academic skills of the children, but differentiation and personalization had a large effect on the probability of exceeding expectations and on the ordinal logit. Differentiation and personalization caused students to be 2.5 more likely to move up a reading level and increased the probability of exceeding expectations by 21.5 percentage points ($p<0.05$).

These results support the previously presented evidence that differentiation is a driver of the results. Any positive feedback mechanism caused by a greater probability of success with the texts is more likely to occur in the tails where differentiation is the greatest. Also possible, however, is the fact that personalization may have been differentially effective for families in the tails of the baseline skill distribution. If parents did not previously realize their child was performing relatively poorly, the new information may have spurred them to more faithfully adhere to the program. Similarly, receiving positive feedback on their child's performance may have encouraged parents at the top of the distribution to build on that success by engaging in the texts to a greater extent.

7. Comparisons with the First Year of the Program

Though there is ample evidence that the personalization and differentiation of the texts provided academic benefits above the general texting curriculum and the control group, the general text messaging curriculum produced no discernable benefits. Table A4 in the online appendix indicates this is the case for both cohorts of participants in the program. At first, this may seem inconsistent with the results from the prekindergarten experiment which estimated that a general texting curriculum can increase preliteracy skills by 10 percent of a standard deviation (York, Loeb, and Doss 2017). A deeper look at the results shows that the first cohort of students in the prekindergarten experiment, which subsequently participated in this follow-up experiment, saw no

significant effect of the general program on preliteracy skills. The general effects on the full population are driven by the second cohort of students that received a combination program.[14] However, both cohorts of the prekindergarten experiment saw large gains for students below the baseline prekindergarten skills distribution. In both years, those children experienced a 30 percent of a standard deviation increase in preliteracy skills. This result provides evidence that general texting programs may be most effective for weaker students.

To probe this question with a direct comparison, we take the students that we followed from the prekindergarten experiment and look at the fall first grade outcomes by baseline of the prekindergarten skills distribution. Sample sizes limit any firm conclusions but the pattern of point estimates indicates that the general texting program may have been more effective for students below the median of skills. Panel A of Table 8 shows that point estimates for the general text condition are large and equal in magnitude to the personalized and differentiated condition. Panel B shows that for students above the median, the point estimates are for the general texting condition are zero or even negative. Again, one must be very cautious in interpreting insignificant point estimates, but the pattern is consistent with the notion that general text messaging curricula benefit weaker students at baseline. Table 7 shows that for students at the tails of the distribution, the effect of personalization and differentiation is largest for the probability of exceeding expectations. This finding further suggests that general texts may be too easy for advanced students, and the differentiation ameliorated this mismatch. This evidence, however, is purely suggestive and we cannot preclude the possibility that elements of the prekindergarten program

[14] We hypothesize that the combined program may impart greater benefits because switching between domains may be more effective at maintaining parents' attention, there may be spillover benefits to literacy by working on math and socio-emotional skills, and success in one domain may encourage parents to continue when they experience difficulties in another domain.

did not replicate. To that end we support replication of the results by providing the text messages used in the prekindergarten experiment and this follow-up study.

8. Robustness Checks

One threat to internal validity of a randomized control trial is differential attrition between treatment and control groups. If different types of people are attriting from each condition our results could be biased. Table 2 presented the overall probability of attrition in each of our models. The probability of attrition is not significantly different in the academic and parent survey samples and marginally significant in the teacher survey sample. We further assessed if there was differential attrition status by covariate and found some imbalance in the teacher survey sample (Table A3 in the online appendix). Because the teacher survey has the greatest amount of attrition, we engage in a Lee (2009) style bounding exercise for that sample of students. Point estimates indicate that fewer people attrited from the two treatment groups. We therefore calculate a trimming proportion, p, for each treatment arm, compared to the control group. We then trim each treatment arm at their respective pth and 1-pth quantile. Re-running our models on these trimmed samples will provide our upper and lower bounds, respectively.

Table 9 presents the results of this bounding exercise. Column 1 and 2 (bolded) present the original estimates from Columns 3 and 4 of Panel C of Table 5. Comparing the original and upper bound estimates, little changed. The effects on the base text treatment remain positive, become slightly larger in magnitude, and become more significant. Coefficients on the differentiated and personalized text messages generally become slightly more positive (or less negative) but their lack of significance remains. In our lower bound estimates, all point estimates become predictably more negative. Point estimates from the general text messaging arm become insignificant and near

zero or slightly negative. Estimates for the differentiated and personalized arm become more negative, and in some cases, significant.

Recall our general conclusion was that there was evidence that the general texting treatment increased parental-teacher contact, but that the differentiation and personalization treatment arm mitigated that effect. The upper bound estimates provide more robust evidence for this inference, while the lower bound estimates indicate that, at worst, the general texts did not affect parent-teacher interaction, and the differentiated and personalized text messages may have significantly decreased interactions. Importantly, the effect of differentiation and personalization relative to the general texting program remains the same in all three estimates. Our overall conclusion therefore remains the same: relative to the general texting program, differentiation and personalization resulted in less parent-teacher contact. This substitution may be due to the greater amount of information contained in the differentiated and personalized texts. The general texts most likely had positive effect on these interactions, though in our most extreme robustness checks they could have had null results.

9. Discussion and Conclusions

In this study, we demonstrate that a low-cost personalized literacy texting intervention can substantially affect student academic outcomes above and beyond a general texting program. Specifically, the differentiation and personalization of the messages caused children to be 63 percent more likely to move up a reading level. Tailoring instruction based on formative assessments has previously been associated with increased student learning in K-12 classrooms (Kulik and Kulik 1984, 1992; Bergen and Sladeczek 1991; Black and William 1998a, 1998b), but this is the first study to show that this approach can also improve parent-child academic interactions. Further, this study provides evidence that text messaging interventions can do more

than merely maintain parents' attention or "nudge" behaviors via reminders. The significant effects of personalization on parent and student outcomes indicate that parents interact with, and absorb the content of the messages as well. This finding supports our hypothesis that the original READY4K! program was effective because it took the complex task of parenting and broke it down into small and easy tasks that were meant to fit into daily life and capitalize on everyday objects.

There are several mechanisms through which the additional gains seen in this study could have been realized. We hypothesized that personalization aspect of the texts could have engendered more trust with the program which would lead to a greater uptake in the activities and thus greater gains in literacy outcomes. Also possible is that information on child performance embedded in the texts updated parental perception of their children's ability and allowed them to more efficiently allocate resources. Meanwhile, the differentiation of the messages helped to better align the difficulty of the task with the child's developmental ability, thus increasing the chance that a parent could successfully engage in the activity with their child. This success may also encourage parents to persist in the program. Finally, the behavior of the parent could have stayed constant and the greater match between text difficulty and student ability could have led to the achievement gains.

Though we cannot definitively pinpoint which mechanisms are at play, survey results indicated that parents in the two treatment text groups saw the texts as equally tailored to their children. This result may indicate that they did not overtly recognize the personalization aspect of the texts, precluding the possibility that personalization engendered more fidelity to the program or that the parents updated their perceptions of their child skill level. More likely, the

differentiation itself led to better outcomes and parents may have had more success in enacting the differentiated activities, leading to a positive feedback loop.

This hypothesis is further supported by our heterogeneity analysis, which indicates that the effects of the differentiation and personalization were particularly concentrated at the tails of the baseline skills distribution where differentiation of the texts was the greatest. There is some evidence that this differentiation was particularly helpful for the higher performing children, as the strongest effects were seen in the probability of exceeding expectations. One may be concerned that the program may increase disparities if the those who were higher at baseline benefit the most from this program. Children in the higher quartiles come from more advantaged families. However, the initial study provided evidence that a general text messaging, with perhaps easier activities, is more effective for children with weaker skills at baseline and the differentiation ameliorated any mismatch between the difficulty of the texts and the skills of the more advanced children. York, Loeb, and Doss (2017) show that the general preliteracy text messaging program had stronger effects for children below the median of the skills distribution. We find suggestive evidence that this occurred in this follow-up experiment as well. The personalization and differentiation may merely improve an intervention that was not previously serving the more academically advanced children in the sample.

The parent and teacher surveys provide additional clues as to how the program changed parent behavior. Parent survey results indicate that recipients of the general text messages thought it was harder to build literacy skills in their children. If the program successfully caused parents to engage in literacy based activities with their child, it is possible that parents realize how hard it is to build academic skills in their children, particularly if the activity and the child's skill level are mismatched. The differentiated and personalized texts successfully mitigated this negative effect,

indicating that the differentiating of the texts may have indeed aligned the child's skill to the activity. The one unanticipated result, however, is that parents in the differentiated and personalized group appear to have substituted away from engaging with teachers.

These results highlight that programs that break down complex tasks, such as building skills in children, can be effective and produce positive outcomes, but that a mismatch between the difficulty of the task and the ability of the parent and child to carry out that task can attenuate any potential gains. Differentiation and personalization of these programs can extract larger gains by minimizing these mismatches, and we demonstrate that even "light-touch" differentiation and personalization based on extant data can generate these gains. The ease and ubiquity of text messaging make it a nimble medium through which educational stakeholders can deliver this differentiated and personalized interventions that minimize frictions caused by mismatch.

Scaling the intervention can be particularly cost effective. The only additional costs over a base texting program are the costs of organizing students into groups according to formative assessment results, the costs of differentiating the activities, and the cost of one sending one extra text message. With the use of current technology we can automate the assignment of students to groups, such that the per-family cost of differentiating the tip tends towards zero as more families are added to the program. The only cost that grows with membership is the texting cost, which is very small compared to other interventions. Overall, differentiating and personalizing text-message interventions based on formative assessment has the promise to produce additional education gains with relatively little additional costs.

References

Allcott, Hunt. 2011. "Social Norms and Energy Conservation." *Journal of Public Economics,* 95(9-10):1082-1095.

Avery, Christopher and Thomas, J Kane. 2004. "Student Perceptions of College Opportunities. The Boston COACH Program." In *College Choices: The Economics of Where to Go, When to Go, and How to Pay For It*, ed. Caroline Hoxby, 355-394. Chicago: The University of Chicago Press.

Bergen, John R., Ingrid E. Sladeczek, and Richard D. Schwarz. 1991. "Effects of A Measurement and Planning System on Kindergarteners' Cognitive Development and Educational Programming." *American Educational Research Journal*, 28(3):683-714.

Bergman, Peter. 2016. "Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment." *Working Paper.* Retrieved April 8, 2016 from: http://www.columbia.edu/~psb2101/BergmanSubmission.pdf

Beshears, John, James J. Choi, David Laibson, Brigitte C. Madrian, Katherine L. Milkman. 2015. "The Effect of Providing Peer Information on Retirement Savings Decisions." *The Journal of Finance,* 70(3):1161-1201.

Black, Paul and Dylan Wiliam. 1998a. "Inside the Black Box: Raising Standards Through Classroom Assessment." *Phi Delta Kappan,* 80(2):139-148.

_____. 1998b. "Assessment and Classroom Learning." *Assessment In Education,* 5(1):7 74.

Cortes, Kalena, Hans Fricke, Susanna Loeb, and David Song. 2017. "Actionable Advice to Foster Parental Engagement: Evidence from An Early-Childhood Text Messaging Experiment." Stanford Mimeo.

Castleman, Benjamin L. and Lindsay C. Page. 2015. "Summer Nudging: Can Personalized Text Messages and Peer Mentor Outreach Increase College Going Among Low-Income High School Graduates?" *Journal of Economic Behavior and Organization, 115*:144-160.

DellaVigna, Stefano. 2009. "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature*, 47(2):315-372.

Dizon-Ross, Rebecca. 2017. "Parents' Perceptions and Children's Education: Experimental Evidence from Malawi." *Working Paper.* Retrieved August 7, 217 from http://faculty.chicagobooth.edu/rebecca.dizon-ross/research/papers/perceptions.pdf

Escueta, Maya, Vincent Quan, Andre Joshua Nickow, and Philip Oreopoulos. 2017. "Education Technology: An Evidence-Based Review." *NBER Working Paper 23744*. Retrieved November 2, 2017 from www.nber.org/papers/w23744.

Frey, Bruno S. and Stephan Meier. 2004. "Social Comparisons and Pro-Social Behavior: Testing 'Conditional Cooperation' in a Field Experiment." *American Economic Review*, 94(5):1717–1722.

Fountas and Pinnell. 2012. "Field Study of Reliability and Validity of the Fountas and Pinnell Benchmark Assessment Systems 1 and 2." Retrieved, July 8, 2015 from http://www.heinemann.com/fountasandpinnell/research/BASFieldStudyFullReport.pdf

Gerber, Alan. S. and Todd Rogers. 2009. "Descriptive Social Norms and Motivation to Vote: Everybody's Voting and So Should You." *The Journal of Politics*, 71(1):178-191.

Grodsky, Eric and Melanie T. Jones. 2007. "Real and Imagined Barriers to College Entry: Perceptions of Cost." *Social Science Research,* 36(2):745-766.

Hastings, Justine S. and Jeffrey M. Weinstein. 2008. "Information, School Choice, and Academic Achievement: Evidence from Two Experiments." *Quarterly Journal of Economics,* 123(4):1373-1414.

Hoxby, Caroline, and Sarah Turner. 2013. "Expanding College Opportunities for High-Achieving, Low Income Students." *Stanford Institute for Economic Policy Research Discussion Paper*, (12-014).

Kast, Felipe, Stephan Meier, and Dina Pomeranz. 2012. "Under-Savers Anonymous: Evidence on Self-Help Groups and Peer Pressure as a Savings Commitment Device." *NBER Working paper 18417*. Retrieved March 30, 2016 from http://www.nber.org/papers/w18417.pdf

Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman. 2016. "Getting To the Top of Mind: How Reminders Increase Saving." *Management Science,* 62(12): 3393-3411.

Kraft, Matthew A. and Todd A. Rogers. 2015. "The Underutilized Potential of Teacher-to-Parent Communication: Evidence from a Field Experiment." *Economics of Education Review,* 47:49-63.

Kulik, Chen-Lin C. and James A. Kulik. 1984. "Effects of Ability Grouping on Elementary School Pupils: A Meta-Analysis." Presented at the 92nd annual meeting of the American Psychological Association in Toronto, Ontario, Canada.

Kulik, James A. and Chen-Lin C. Kulik. 1992. "Meta-analytic findings on grouping programs." *Gifted Child Quarterly,* 36(2):73-77.

Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies,* 76(3):1071-1102.

Lou, Yipling, Philip C. Abrami, John C. Spence, Catherine Poulsen, Bette Chambers, and Sylvia d'Apollinia. 1996. "Within-class grouping: A meta-analysis." *Review of Educational Research*, 66(4):423-458.

Mullainathan, Sendhil, and Richard H. Thaler. 2000. "Behavioral Economics." *NBER Working Paper No. w7948*. Retrieved December 15, 2015 from http://www.nber.org/papers/w7948

Pane, John.F, Beth Ann Griffin, Daniel F. McCaffrey, and Rita Karam. 2014. "Effectiveness of Cognitive Tutor Algebra I at Scale." *Educational Evaluation and Policy Analysis,* 36(2):127-144.

Robinson, Joseph P. 2009. "Evidence of a Differential Effect of Ability-Grouping on the Reading Achievement Growth of Language-Minority Hispanics." *Educational Evaluation and Policy Analysis*, 30(2):141-180.

Samejima, Fumiko. 1997. "Graded Response Model." In *Handbook of Modern Item Response Theory*, eds. Wim. J. van der Linden and Ronald K. Hambleton, 85–100. New York, NY: Springer New York.

Valant, Jon, and Susanna Loeb. 2014. "Information, Choice, and Decision-Making: Field Experiments with Adult and Student School Choosers." *Working Paper*.

Van Klaveren, Chris, Sebastiaan Vonk, and Ilja Cornelisz. 2017. "The Effect of Adaptive Versus Static Practicing on student Learning- Evidence from a Randomized Field Experiment." *Economics of Education Review,* 58:175-187.

Vygotsky, Lev S. 1978a. "Interaction between learning and development." *Readings on the Development of Children,* 23(3):34-41.

_____. 1978b. "Mind in Society: The Development of Higher Psychological." Cambridge, MA: Harvard University.

York, Benjamin N., Susanna Loeb, and Christopher Doss. 2017. "One Step at a Time: The Effect of an Early Literacy Text Messaging Program for Parents of Preschoolers." *NBER Working Paper 20659*. Retrieved December 2, 2014 from: http://www.nber.org/papers/w20659

	General Example 1	Personalized Example 1			
		Quartile 1	Quartile 2	Quartile 3	Quartile 4
Monday	FACT: Beginning word sounds are often made up of multiple letters like “th” or “st”. Learning these sounds is a key to reading.				
Wednesday	TIP: As your child gets dressed ask: what sound does SHOE start with? What letters are in ‘shh’? (s and h) What else starts with ‘shh’? Shirt!	TIP: Your child’s fall K test shows s/he is starting to learn beginning word sounds. Support this progress with simple activities! As your child gets dressed say: Shhh-oe starts with shhh. Do you know what else starts with shhh? Shh-irt!! What letters are in shhh? (s and h)	TIP: Your child’s fall K test shows his/her knowledge of beginning word sounds is growing. Support this progress with simple activities! As your child gets dressed ask: what sound does SHOE start with? What letters are in ‘shh’? (s and h) What else starts with ‘shh’? (Shirt)	TIP: Your child’s fall K test shows his/her knowledge of beginning word sounds is solid. Support this progress with simple activities! As your child gets dressed say: What are 2 things you wear that start with the ‘shhh’ sound? (Shoes and Shirt) What letters are in shhh? (s and h)	TIP: Your child’s fall K test shows his/her knowledge of beginning word sounds is strong. Support this progress with simple activities! As your child gets dressed say: Name things we wear that start with the ‘shhh’ sound. (shoes shorts shirt) What letters are in shhh? (s and h)
Friday	GROWTH: Keep practicing word sounds! Now ask: what sound does brrr-eakfast start with? (Brrr) What foods start with brrr? (Bread, brownie)	GROWTH: Keep practicing word sounds! Ask: what sound does brr-eakfast start with? (Brrr) Name a food that starts with brrr (Bread)	GROWTH: Keep practicing word sounds! Ask: what sound does brrr-eakfast start with? (Brrr) What 2 foods start with brr? (Bread, brownie)	GROWTH: Keep practicing word sounds! Ask: what sound does brr-eakfast start with? What foods start with brr? (Bread, brownie, broccoli)	GROWTH: Keep practicing word sounds! Say: Name as many foods as you can that start with the same sound as brrr-eakfast (Bread, brownie)

	Control Example 1	Control Example 2
Wednesday	TIP: Planning for school emergencies is important. Make sure that you filled out the Emergency Card and returned it to the school office.	TIP: SFUSD is all about great food. Did you know local chefs hand prepare our meals fresh daily? Go to www.sfusd.edu to learn more.

Figure 1: Text Examples

Table 1
Descriptive Statistics

	Parent Survey Sample (N=475)		Teacher Survey Sample (N = 409)		Academic Outcomes Sample (N=578)		SFUSD Kindergarten Cohort (N=4,532)	
Variable	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Panel A: Children								
Male	0.520	0.500	0.531	0.500	0.509	0.500	0.509	0.500
Hispanic	0.324	0.469	0.333	0.472	0.346	0.476	0.266	0.442
Asian	0.326	0.469	0.342	0.475	0.334	0.472	0.234	0.426
Decline To State Ethnicity	0.046	0.210	0.056	0.231	0.057	0.232	0.166	0.372
White	0.074	0.262	0.071	0.257	0.074	0.263	0.144	0.651
Other	0.168	0.375	0.198	0.399	0.189	0.392	0.188	0.391
Age in Years	5.423	0.294	5.458	0.292	5.431	0.297	5.497	0.297
Enrolled in Transitional Kindergarten	0.053	0.224	0.049	0.216	0.066	0.248	0.082	0.275
Parent rating of letter knowledge	3.055	0.903	3.054	0.903	3.045	0.917		
Parent rating of letter sounds	3.191	1.142	3.177	1.179	3.214	1.173		
Parent rating of rhyming	3.051	1.225	3.051	1.255	3.031	1.242		
Panel B: Parents								
Has less than a bachelor's degree	0.636	0.482	0.636	0.482	0.651	0.477		
Received Texts in English	0.543	0.499	0.523	0.500	0.526	0.500		
Received Texts in Spanish	0.242	0.429	0.247	0.432	0.260	0.439		
Received Texts in Chinese	0.215	0.411	0.230	0.421	0.215	0.411		
First Year Receiving Texts	0.383	0.487	0.457	0.499	0.422	0.494		
Age in Years	34.90	6.185	34.347	6.129	34.38	6.248		
Times per week read to child	2.979	0.864	2.913	0.877	2.952	0.871		
Times per week told stories to child	2.792	0.868	2.761	0.846	2.769	0.856		
Times per week sang to child	2.856	0.862	2.837	0.850	2.822	0.845		
Times per week child ask to be read to	2.943	0.914	2.914	0.908	2.941	0.907		

Note: Parents rated the letter knowledge of their child in one of four categories: 1=The child knows no letters, 2=Some, 3=Most, 4=All. Parents rated how well their child can produce letter sounds and rhyme in one of five categories: 1=Not at all, 2=Not very well, 3=Somewhat well, 4=Well, 5=Very Well. Answer options for weekly parental activities and how often the child asks to be read to include: 1=Not at all, 2=Once or twice per week, 3=Three to six times, 4=Every day. Missing values set at the sample average. For families in first year of experiment the baseline survey questions were given in September 2014. For families in the second year of the experiment the baseline survey questions were given in September 2013. All child demographics are from San Francisco Unified School District administration files.

Table 2
Overall Attrition

	(1) General Text Treatment	(2) Personalized Text Treatment
Not in Parent Survey Sample	0.003	-0.007
	(0.0501)	(0.0496)
Not in Teacher Survey Sample	-0.066	-0.075+
	(0.0398)	(0.0378)
Not in Academic Sample	-0.016	-0.0189
	(0.0389)	(0.0402)

Note: Each pair of cells represents the results of a separate regression of the treatment effect on an indicator for not being in the sample defined by the row header. Column headers indicate the model components. N = 794. Models include randomization site fixed effects. Standard errors are clustered by randomization site. Parent survey sample refers to end line parent survey respondents, teacher survey sample refers to end line teacher survey respondents, and academic sample refers to children tested in fall of first grade on the Fountas and Pinnell Benchmark Assessment system. +indicates $p<0.10$, *$p<0.05$, **$p<0.01$, ***$p<0.001$

Table 3
Effects on Fountas and Pinnell Benchmark Assessment System academic outcomes

	(1) General Text Treatment	(2) Personalized Text Treatment	(3) General Text Treatment	(4) Personalized Text Treatment
Panel A: Reading Level Outcomes				
Reading level (ordinal logit)	-0.0638	0.2753	0.0133	0.4915***
	(0.1478)	(0.1811)	(0.1769)	(0.1331)
Reading level(standardized point scale)	0.0031	0.1851+	0.0058	0.1828*
	(0.0804)	(0.0948)	(0.0779)	(0.0782)
Pr(Reading Level A or Above)	-0.0037	0.0264	0.0013	0.0177
	(0.0253)	(0.0337)	(0.0236)	(0.0314)
Pr(Reading Level C or Above)	0.0115	-0.0035	0.0162	-0.0086
	(0.0417)	(0.0494)	(0.0383)	(0.0408)
Pr(Reading Level E or Above)	-0.0027	0.0890+	0.0085	0.0884*
	(0.0454)	(0.0467)	(0.0453)	(0.0385)
Pr(Reading Level G or Above)	-0.0061	0.0413	-0.0101	0.0414
	(0.0437)	(0.0414)	(0.0444)	(0.0388)
Panel B: District Academic Benchmarks,				
Exceeds Expectations	-0.0127	0.1180**	-0.013	0.1205***
	(0.0466)	(0.0380)	(0.0449)	(0.0331)
Meets or Exceeds Expectations	-0.0052	0.0869*	0.0037	0.0900**
	(0.0411)	(0.0368)	(0.0373)	(0.0301)
Approaches, Meets, or Exceeds Expectations	0.0263	0.0094	0.0291	0.0074
	(0.0508)	(0.0466)	(0.0474)	(0.0410)
Randomization Site Fixed Effects	✓	✓	✓	✓
Language of Texts	✓	✓	✓	✓
Baseline Survey Controls			✓	✓
Administrative Covariates			✓	✓

Note: Each pair of cells represents the results of a separate regression of the treatment effect on the relevant academic outcome. Column headers indicate the model components. Row headers indicate the academic outcome. A Graded Response Model was used to create the factors of baseline survey responses. Factors were made from parent reports of parent age and education, parent reports of the child's knowledge of letters, letter sounds, and rhyming, parents reports of the frequency with which the parent read to, told stories to, and sang to their child, and parent reports of how often the child asks questions. N = 578 for all regressions. Source data are district test files of the Fountas and Pinnell Benchmark Assessment System in fall of first grade. Standard errors are clustered at the randomization site level. +indicates p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table 4
Effects on parent beliefs and home activities

	(1) General Text Treatment	(2) Personalized Text Treatment	(3) General Text Treatment	(4) Personalized Text Treatment	N
Panel A: Parent Outcome Factors					
Parent Belief Factor	-0.1459	0.0539	-0.1551	0.031	455
	(0.1395)	(0.1444)	(0.1445)	(0.1339)	
Activities Factor	0.0856	0.2717*	0.1375	0.2628*	421
	(0.1301)	(0.1020)	(0.1314)	(0.1056)	
Panel B Parent Beliefs					
Enjoys home literacy activities	-0.1514	-0.0369	-0.1572	-0.0359	462
	(0.1127)	(0.1397)	(0.1157)	(0.1297)	
Knows literacy skills needed for first grade	-0.1361	0.0908	-0.1558	0.0525	458
	(0.1348)	(0.1355)	(0.1381)	(0.1224)	
Believes can build literacy skills	-0.1357	0.0165	-0.1516	-0.0152	459
	(0.1328)	(0.1453)	(0.1395)	(0.1374)	
Believes he/she plays an important role in building literacy skills	-0.0613	-0.0536	-0.0575	-0.0542	460
	(0.1248)	(0.1391)	(0.1302)	(0.1262)	
Building reading skills is easy	-0.2659+	0.0854	-0.2737+	0.0487	459
	(0.1447)	(0.1409)	(0.1431)	(0.1418)	
Feels supported	0.0132	0.2114+	0.014	0.2014	457
	(0.1508)	(0.1256)	(0.1578)	(0.1222)	
Panel C: Parent Activities					
Read words with child	0.2371*	0.2022	0.2423*	0.1653	463
	(0.1175)	(0.1233)	(0.1205)	(0.1231)	
Wrote notes with child	0.0581	0.1239	0.0618	0.0976	461
	(0.1276)	(0.1432)	(0.1326)	(0.1436)	
Took books when left the house	0.2183+	0.2720+	0.2674*	0.2829+	461
	(0.1147)	(0.1395)	(0.1065)	(0.1452)	
Read books to child	0.0158	0.1374	0.053	0.1426	454
	(0.1129)	(0.1052)	(0.1103)	(0.1010)	
Had child read books to parent	0.0609	-0.0364	0.0891	-0.0677	461
	(0.1510)	(0.1136)	(0.1296)	(0.1038)	
Reviewed parts of a book	0.0101	0.2191*	0.0398	0.2102+	462
	(0.1277)	(0.1004)	(0.1266)	(0.1092)	
Reviewed direction of reading	-0.0017	0.3336***	0.0084	0.3154**	461
	(0.1355)	(0.0857)	(0.1324)	(0.0950)	
Corrected mistakes while reading	0.1167	0.2794*	0.1227	0.2502*	462
	(0.1150)	(0.1065)	(0.1241)	(0.1083)	
Asked child questions about book	0.0564	0.1652	0.0938	0.1767	464
	(0.1199)	(0.1234)	(0.1236)	(0.1268)	
Practiced rhyming	0.144	0.3535***	0.1759	0.3699***	460
	(0.1515)	(0.1003)	(0.1429)	(0.0910)	
Practiced writing child's name	0.0879	0.1604	0.1164	0.151	459
	(0.1336)	(0.1268)	(0.1333)	(0.1387)	
Randomization Site Fixed Effects	✓	✓	✓	✓	
Language of Texts	✓	✓	✓	✓	
Baseline Survey Controls			✓	✓	
Administrative Covariates			✓	✓	

Note: Each pair of cells represents the results of a separate regression of the treatment effect on the relevant outcome. Column headers indicate the model components. Row headers indicate the literacy outcome. All literacy outcomes are standardized. Factor analysis was used to determine the outcome factors. See Table A1 in the online appendix for a list of survey questions that compose each outcome factor. Covariates are detailed in Table 3. Standard errors are clustered at the randomization site level. Source data are the responses to parent surveys. +indicates $p<0.10$, *$p<0.05$, **$p<0.01$, ***$p<0.001$.

Table 5
Effects on parent interactions with teachers

	(1) General Text Treatment	(2) Personalized Text Treatment	(3) General Text Treatment	(4) Personalized Text Treatment	N
Panel A: Outcome Factors					
Parent Report of Interactions Factor (N = 388)	0.2150+ (0.1271)	0.1595+ (0.0892)	0.2295+ (0.1209)	0.1136 (0.0812)	472
Teacher Report of Interactions Factor (N = 348)	0.1843 (0.1338)	-0.218 (0.1495)	0.2162 (0.1333)	-0.1739 (0.1537)	368
Panel B: Parent Reports on Interactions with Teacher					
Talked to teacher	0.1917 (0.1278)	0.0744 (0.1110)	0.2060+ (0.1204)	0.0499 (0.1022)	475
Talked to teacher about child's interests	0.0765 (0.1201)	0.0983 (0.1010)	0.0759 (0.1200)	0.036 (0.1033)	474
Talked to teacher about how well child is getting along with others	0.2460+ (0.1238)	0.1431+ (0.0810)	0.2658* (0.1261)	0.1126 (0.0886)	475
Talked to teacher about how well child is doing in school	0.1485 (0.1291)	0.2102* (0.0997)	0.1593 (0.1196)	0.1657+ (0.0863)	473
Talked to teacher about child's early literacy skills	0.1104 (0.1284)	0.1697+ (0.0896)	0.1094 (0.1258)	0.1259 (0.0888)	474
Talked to teacher about child's reading skills	0.1977 (0.1339)	0.1441 (0.1001)	0.1878 (0.1267)	0.0974 (0.0926)	474
Asked for book and home activity recommendations	0.184 (0.1233)	0.1448 (0.1232)	0.2014+ (0.1141)	0.1052 (0.1231)	474
How well does parent know teacher	0.2319+ (0.1195)	0.0533 (0.0975)	0.2592* (0.1128)	0.0455 (0.0840)	474
Panel C: Teacher Reports on Interactions with Parents					
Parent talks about child's interests	0.1940+ (0.1132)	-0.1446 (0.1487)	0.2312+ (0.1185)	-0.0892 (0.1528)	396
Parent asks how well child gets along with others	0.161 (0.1520)	-0.0814 (0.1595)	0.1982 (0.1580)	-0.03 (0.1727)	395
Parent asks how well child is doing in school	0.165 (0.1626)	-0.1961 (0.1792)	0.1982 (0.1585)	-0.1357 (0.1901)	397
Parent asks about child's literacy skills	0.196 (0.1289)	-0.2259* (0.1000)	0.2553+ (0.1321)	-0.1751 (0.1128)	396
Parent asks how to help child learn to read	0.1912 (0.1442)	-0.1905 (0.1540)	0.2063 (0.1490)	-0.1775 (0.1579)	397
Parent asks for book recommendations	0.0227 (0.1160)	-0.1683 (0.1267)	0.0251 (0.1273)	-0.1528 (0.1438)	396
Parent talks about home activities	0.2825* (0.1216)	-0.0759 (0.1484)	0.2787* (0.1131)	-0.0571 (0.1720)	409
How well does teacher know parent	-0.0315 (0.1575)	-0.2453+ (0.1378)	-0.0762 (0.1709)	-0.2610+ (0.1527)	382
Randomization Site Fixed Effects	✓	✓	✓	✓	
Language of Texts	✓	✓	✓	✓	
Factors of Baseline Survey Responses			✓	✓	
Administrative Covariates			✓	✓	

Note: Each pair of cells represents the results of a separate regression of the treatment effect on the relevant outcome. Column headers indicate the model components. All outcomes are standardized. Factor analysis was used to create the outcome factors. The parent report of interactions factor is made up of the questions in Panel A, the teacher report of interactions factor is made up questions in Panel B. Covariates are detailed in Table 3. Source data in Panels A and B are the responses to parent surveys fielded after the texting program ended in August of 2015. Source data in Panels A and C are the responses to teacher surveys fielded in May of 2015. +indicates p<0.10, *p<0.05, **p<0.01, ***p<0.001.

Table 6
Effects on parental attitudes towards texts

	(1)	(2)	(3)	(4)	
	General Text Treatment	Personalized Text Treatment	General Text Treatment	Personalized Text Treatment	N
Read Texts	0.0101	-0.0154	0.0503	-0.0184	474
	(0.1105)	(0.1333)	(0.1068)	(0.1242)	
Used Texts	0.1603	0.2781*	0.1646	0.2879*	474
	(0.1082)	(0.1156)	(0.1079)	(0.1134)	
Texts were helpful	0.3129**	0.4601***	0.3066**	0.4560***	473
	(0.0911)	(0.1014)	(0.0912)	(0.0988)	
Texts were made for you	0.3982***	0.4324***	0.3857***	0.4529***	473
	(0.0921)	(0.0996)	(0.1020)	(0.1046)	
Would recommend texts to others	0.1019	0.2475*	0.073	0.2292*	471
	(0.1268)	(0.1018)	(0.1291)	(0.1053)	
Randomization Site Fixed Effects	✓	✓	✓	✓	
Language of Texts	✓	✓	✓	✓	
Factors of Baseline Survey Responses			✓	✓	
Administrative Covariates			✓	✓	

Note: Each pair of cells represents the results of a separate regression of the treatment effect on the relevant outcome. Column headers indicate the model components. All outcomes are standardized. Covariates are detailed in Table 3. Standard errors are clustered at the randomization site level. Source data are the responses to parent surveys fielded after the texting program ended in August of 2015. +indicates p<0.10, *p<0.05, **p<0.01, ***p<0.001.

Table 7
Heterogeneity of academic outcomes by baseline academic performance (first year of experiment only)

	(1)	(2)
	General Text Treatment	Personalized Text Treatment
Panel A: Quartiles 2 and 3 (N = 123)		
Reading level (ordinal logit)	0.3545	0.4253
	(0.3405)	(0.3780)
Reading level (standardized point scale)	0.0645	0.0895
	(0.1095)	(0.0817)
Exceeds Expectations	0.0092	0.0495
	(0.1089)	(0.0799)
Meets or Exceeds Expectations	0.1713+	0.2196
	(0.0882)	(0.1346)
Approaches, Meets, or Exceeds Expectations	0.051	0.046
	(0.1015)	(0.1515)
Panel B: Quartiles 1 and 4 (N = 121)		
Reading level (ordinal logit)	-0.3068	0.9128*
	(0.7124)	(0.4183)
Reading level (standardized point scale)	-0.1222	0.3631
	(0.2898)	(0.2348)
Exceeds Expectations	-0.0722	0.2154*
	(0.1048)	(0.0844)
Meets or Exceeds Expectations	-0.0934	0.0728
	(0.1169)	(0.0824)
Approaches, Meets, or Exceeds Expectations	-0.0559	0.0361
	(0.1258)	(0.0795)

Note: Each pair of cells represents the results of a separate regression of the treatment effect on the relevant academic outcome. Column headers indicate the model components. The reference category is the control group. Row headers indicate the academic outcome. Panel headers indicate the quartile subsample. All models include randomization site fixed effects, controls for texting language, factors of baseline survey responses, and administrative covariates. Covariates are detailed in Table 3. Source data are district test files of the Fountas and Pinnell Benchmark Assessment System in fall of first grade. Standard errors are clustered at the randomization site level. +indicates $p<0.10$, *$p<0.05$, **$p<0.01$, ***$p<0.001$

Table 8
Effects academic skills on second year participants, by median of prekindergarten baseline skills distribution

	Model 1	
	(1)	(2)
	General Text Treatment	Personalized Text Treatment
Panel A: Below Median of Baseline Skills (N=165)		
Reading level (ordinal logit)	0.5437	0.6644
	(0.6224)	(0.4397)
Reading level (standardized point scale)	0.1414	0.1426
	(0.1316)	(0.1253)
Exceeds Expectations	0.0739	0.1628*
	(0.0764)	(0.0735)
Meets or Exceeds Expectations	0.148	0.148
	(0.1012)	(0.1012)
Approaches, Meets, or Exceeds Expectations	0.1727	0.0261
	(0.1415)	(0.1114)
Panel B: Above Median of Baseline Skills (N=166)		
Reading level (ordinal logit)	-0.6473	0.4734
	(0.5700)	(0.3280)
Reading level (standardized point scale)	-0.0168	0.2114
	(0.2394)	(0.2254)
Exceeds Expectations	0.0045	0.0645
	(0.1117)	(0.0726)
Meets or Exceeds Expectations	-0.1269	0.0714
	(0.0909)	(0.0664)
Approaches, Meets, or Exceeds Expectations	0.011	0.0739
	(0.0847)	(0.0555)

Note: Each pair of cells represents the results of a separate regression of the treatment effect on the relevant academic outcome. Column headers indicate the model components. The reference category is the control group. Row headers indicate the academic outcome. Panel headers indicate the subsample. All models include the covariates detailed in Table 3 including randomization site fixed effects, control for text message language, factors of baseline survey questions, and administrative covariates. Source data are district test files of the Fountas and Pinnell Benchmark Assessment System in fall of first grade. Baseline skills were calculated from performance on the Phonological Awareness Literacy Screening administered in fall of 2013. Standard errors are clustered by randomization site. +indicates p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table 9
Bounds on teacher survey outcomes

	Original Estimates		Upper Bounds		Lower Bounds			
	(1)	(2)	(3)	(4)	(5)	(6)		
	General Text Treatment	Personalized Text Treatment	General Text Treatment	Personalized Text Treatment	General Text Treatment	Personalized Text Treatment	N(Original)	N(Lee Bounds)
Teacher Report of Interactions Factor	**0.2162**	**-0.1739**	0.3313*	-0.0018	0.0036	-0.1687	368	339
	(0.1333)	**(0.1537)**	(0.1264)	(0.1650)	(0.1315)	(0.1909)		
Parent talks about child's interests	**0.2312+**	**-0.0892**	0.3666*	0.0678	0.0584	-0.2497	396	367
	(0.1185)	**(0.1528)**	(0.1765)	(0.1866)	(0.0839)	(0.1617)		
Parent asks how well child gets along with others	**0.1982**	**-0.03**	0.3729*	0.1825	0.0859	-0.2598	395	367
	(0.1580)	**(0.1727)**	(0.1730)	(0.1943)	(0.1663)	(0.1724)		
Parent asks how well child is doing in school	**0.1982**	**-0.1357**	0.2412	0.0965	0.0326	-0.2593	397	367
	(0.1585)	**(0.1901)**	(0.1629)	(0.1907)	(0.1765)	(0.1866)		
Parent asks about child's literacy skills	**0.2553+**	**-0.1751**	0.3531*	-0.0255	-0.0282	-0.3225**	396	367
	(0.1321)	**(0.1128)**	(0.1395)	(0.1159)	(0.1391)	(0.1060)		
Parent asks how to help child learn to read	**0.2063**	**-0.1775**	0.3770*	0.0295	-0.0992	-0.4491**	397	367
	(0.1490)	**(0.1579)**	(0.1594)	(0.1728)	(0.1303)	(0.1310)		
Parent asks for book recommendations	**0.0251**	**-0.1528**	0.0993	-0.0574	-0.1815	-0.3850**	396	367
	(0.1273)	**(0.1438)**	(0.1538)	(0.1705)	(0.1316)	(0.1371)		
Parent talks about home activities	**0.2787***	**-0.0571**	0.3770**	0.1502	0.1132	-0.2381	409	379
	(0.1131)	**(0.1720)**	(0.1182)	(0.1535)	(0.1067)	(0.1882)		
How well does teacher know parent	**-0.0762**	**-0.2610+**	0.0959	-0.0827	-0.2474	-0.3581*	382	351
	(0.1709)	**(0.1527)**	(0.1671)	(0.1851)	(0.1805)	(0.1659)		

Note: Each pair of cells represents the results of a separate regression of the treatment effect on the relevant outcome. Row headers indicate the outcome. Column headers indicate model components. All outcomes are standardized. Factor analysis was used to create the outcome factors. Upper and lower bound estimates were calculated by the procedure recommended by Lee (2009). All models include randomization site fixed effects, controls for texting language, factors of baseline survey responses, and administrative covariates. Covariates are detailed in Table 3. Source data in Panels A and C are the responses to teacher surveys fielded in May of 2015. Standard errors are clustered at the randomization site level. +indicates p<0.10, *p<0.05, **p<0.01, ***p<0.001.

Appendix

	General Example 2	**Personalized Example 2**			
		Quartile 1	Quartile 2	Quartile 3	Quartile 4
Monday	FACT: Spelling can be a fun way for your child to practice his/her reading and writing skills in a new way!				
Wednesday	TIP: Say: Let's spell the word "he". Sound it out. What makes the "hhh" sound? What makes the "eee" sound? "He" is spelled H-E. Try we and be.	TIP: Here is a tip based on your child's K literacy exam. Say: "Let's spell the word 'he'. Sound it out. 'H' makes the 'hhh' sound. 'E' makes the 'eee' sound 'He' is spelled H-E." Do it again with we and be.	TIP: Here is a tip based on your child's K literacy exam. Say: "Let's spell the word 'he'. Sound it out. What makes the 'hhh' sound? What makes the 'eee' sound? 'He' is spelled H-E. Now you try to spell we and be."	TIP: Here is a tip based on your child's K literacy exam. Say: "Let's spell the word 'he'. Sound it out. What makes the 'hhh' sound? What makes the 'eee' sound? 'He' is spelled H-E. What rhymes with 'he' (we, be, she). Can you spell those words?"	TIP: Here is a tip based on your child's K literacy exam. Say: "Let's spell the word 'he'. Sound it out. What makes the 'hhh' sound? What makes the 'eee' sound? 'He' is spelled H-E. What rhymes with 'he' (we, be, she). Can you spell those words? Can you write them down?"
Friday	GROWTH: Keep spelling! Have a spelling bee at home. First you spell a word (my, is, no). Then ask your child to spell one (by, it, go).	GROWTH: Keep spelling! Now ask your child spell words like 'my', 'by', and 'shy'.	GROWTH: Keep spelling! Have a spelling bee at home. First you spell a word (my, is, no). Then ask your child to spell one (by, it, go).	GROWTH: Keep spelling! Have a spelling bee. You spell a word (my, is, no). Then your child spells one (by, it, go). Take turns writing them down.	GROWTH: Keep spelling! Have a spelling bee. You spell a word (my/no). Then your child spells a rhyming word (by/go). Take turns writing them.

Figure A1: Additional Text Examples

Table A1
Factor analysis details

Component	Scoring Coefficient
Panel A: Parent belief factor (N=455)	
I enjoy doing activities with my child that build his/her reading skills	0.19260
I know which literacy skills my child needs to be ready for first grade	0.20657
I know what I can do to help my child build the literacy skills necessary for 1st	0.20496
I play an important role in building my child's reading skills	0.19675
Building my child's reading skills is easy	0.17396
I feel supported in helping prepare my child for first grade	0.20099
Eigenvalue: 4.32627 (72.10% of variance explained)	
Panel B: Activities factor (N=421)	
Last week, how many times did you do each of the following reading related activities with your child?	
Practiced reading words	0.06533
Write a note to you child for him/her to read	0.06543
Brought books when leaving the house	0.05682
Read to your child	0.06038
Had your child read to you	0.07063
Showed your child the different parts of a book	0.06847
Showed your child that we read from left to right	0.06640
Asked your child to follow the words with her/her finger as your read	0.07554
Asked questions about the pictures in a book	0.07843
Worked with your child to correct his/her mistakes as s/he read	0.06746
Asked your child questions about a book or story s/he recently read or heard	0.07513
Last week how many times did you do each of the following literacy skills activities with your child?	
Said beginning word sounds with your child	0.07149
Hunted for lower and upper case letters in a book or magazine	0.07216
Said a new word to your child and talked about what it means	0.07467
Asked your child questions to spark his/her imagination	0.06705
Said ending word sounds with your child	0.07582
Hunted for small words in a book or magazine	0.07155
Said rhyming words with your child	0.07792
Helped your child write his/her name	0.05825
Had your child describe the things s/he sees	0.07235
Had your child help you with a daily routine	0.06063
Played a game with your child like "I spy"	0.06951
Eigenvalue: 9.43160 (42.87% of variance explained)	
Panel C: Parent reports of school involvement factor (N=472)	
During a typical week, how many times did you talk to your child's teacher	0.12826
How well did you know your child's kindergarten teacher	0.12096
Since January, how times did you talk to your child's teacher about:	
Your child's interests	0.16105
How your child is getting along with other children	0.16008
How your child is doing in school	0.16802
Your child's early literacy skills	0.16666
Your child's reading skills	0.17171
Books that your child might like or activities to do at home with your child	0.16563
Eigenvalue: 5.11333 (63.92% of variance explained)	
Panel D: Teacher reports of parent involvement factor (N=348)	
How well do you know the parents of (child's name)	0.12023
How often do parents of (child's name) ask you about the following topics	
Their child's interests	0.16266
How their child gets along with others	0.16110
What their child is doing in school	0.17707
Their child's understanding of early literacy skills like letter sounds	0.17832
Things they can do to help their child learn to read	0.18057
Book recommendations	0.14672
Tell you about what they are doing at home to help their child learn	0.17137
Eigenvalue: 4.68350 (58.54% of variance explained)	

Table A2
Covariate Balance

	Parent Survey Sample (N=475)		Teacher Survey Sample (N=409)		Academic Sample (N=578)	
	(1)	(2)	(3)	(4)	(5)	(6)
	General Text Treatment	Personalized Text Treatment	General Text Treatment	Personalized Text Treatment	General Text Treatment	Personalized Text Treatment
Panel A: Factors of Baseline Survey Questions						
Literacy Skills Factor	-0.0776	-0.1487	-0.004	-0.2334	-0.054	-0.1707
	(0.1457)	(0.1424)	(0.1669)	(0.1791)	(0.1078)	(0.1175)
Home Activities Factor	0.0.278	0.0625	0.1691	-0.0129	0.0332	-0.0201
	(0.0988)	(0.0764)	(0.1039)	(0.0954)	(0.0759)	(0.0843)
Parent Background Factor	0.1448	0.1419	0.1741	0.0117	0.1175	0.0694
	(0.1465)	(0.1760)	(0.1639)	(0.2079)	(0.1383)	(0.1744)
Panel B: Child Covariates						
Male	-0.014	-0.026	-0.0307	-0.0888+	-0.0204	-0.0459
	(0.0490)	(0.0519)	(0.0608)	(0.0526)	(0.0445)	(0.0480)
Hispanic	-0.0006	-0.0212	-0.0212	-0.0009	0.0086	-0.0123
	(0.0538)	(0.0457)	(0.0509)	(0.0400)	(0.0379)	(0.0360)
Asian	0.0144	0.0075	0.0747	0.021	0.0416	-0.0008
	(0.0458)	(0.0475)	(0.0495)	(0.0564)	(0.0324)	(0.0367)
Decline To State Ethnicity	-0.0162	-0.0065	-0.0062	-0.0135	-0.0048	-0.0132
	(0.0245)	(0.0243)	(0.0254)	(0.0332)	(0.0234)	(0.0268)
White	-0.0476	-0.0267	-0.0504*	-0.0345	-0.0677**	-0.035
	(0.0297)	(0.0252)	(0.0213)	(0.0332)	(0.0237)	(0.0231)
Other	0.0349	0.0569	0.0031	0.0279	0.0223	0.0613
	(0.0461)	(0.0488)	(0.0553)	(0.0496)	(0.0377)	(0.0414)
Age in Years	-0.0078	0.057	0.0281	0.0653	-0.017	0.0262
	(0.0349)	(0.0365)	(0.0349)	(0.0360)	(0.0266)	(0.0346)
Enrolled in Transitional Kindergarten	-0.0042	0.0312	-0.0116	0.0329	0.0202	0.0305
	(0.0167)	(0.0375)	(0.0252)	(0.0267)	(0.0226)	(0.0351)
Panel C: Parent Covariates						
Received Texts in English	0.0544	-0.0024	0.0131	-0.0148	0.0375	0.0174
	(0.0338)	(0.0324)	(0.0289)	(0.0325)	(0.0307)	(0.0290)
Received Texts in Spanish	-0.0312	-0.0261	-0.0091	-0.0083	-0.021	-0.0165
	(0.0318)	(0.0296)	(0.0267)	(0.0319)	(0.0227)	(0.0277)
Received Texts in Chinese	-0.0232	0.0285	-0.004	0.0231	-0.0164	-0.0009
	(0.0237)	(0.0291)	(0.0227)	(0.0282)	(0.0207)	(0.0292)

Note: Each pair of cells represents the results of a separate regression of the treatment effect on the relevant covariate. Column headers indicate the sample and model components. The reference category is the control group. Row headers indicate the covariate tested. A Graded Response Model was used to create the covariate factors. The literary skills factor was made from the parent ratings of the child's letter knowledge, letter sounds, and rhyming. The home activities factor was made from parent reports of how often they read, told stories, and sang with their child, and how often the child asked to be read to. The parent background factor was made with parent age and education. All regressions include randomization site fixed effects. Standard errors are clustered at the randomization site level. Source data in Panels A and C are the responses to parent surveys fielded in August of 2015. Source data in Panel B are San Francisco Unified School District Administrative Files. +indicates p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table A3
Attrition by student characteristic

	(1)	(2)
	General Text Treatment * Covariate	Personalized Text Treatment * Covariate
Panel A: Not in Parent Survey Sample		
English	-0.1199	-0.0832
	(0.0909)	(0.0743)
Spanish	0.0827	0.1399
	(0.1118)	(0.0971)
Chinese	0.0945	-0.0385
	(0.1140)	(0.1141)
Male	-0.0042	0.0733
	(0.0756)	(0.0897)
Age in Years	-0.037	-0.0642
	(0.1185)	(0.1105)
First Year in Program	-0.0306	0.0258
	(0.1100)	(0.1019)
Panel B: Not in Teacher Survey Sample		
English	-0.0292	0.0038
	(0.0739)	(0.0659)
Spanish	0.08	-0.0044
	(0.0939)	(0.0774)
Chinese	-0.0558	0.0087
	(0.0828)	(0.0885)
Male	0.0296	0.0619
	(0.0831)	(0.0773)
Age in Years	-0.2297+	-0.0677
	(0.1155)	(0.1116)
First Year in Program	0.0988	0.0705
	(0.0771)	(0.0650)
Panel C: Not in Academic Sample		
English	-0.0153	-0.0884
	(0.0756)	(0.0716)
Spanish	0.0291	0.0316
	(0.0939)	(0.1055)
Chinese	-0.0156	0.106
	(0.0671)	(0.1244)
Male	-0.0528	0.0432
	(0.0801)	(0.0766)
Age in Years	0.0475	0.0255
	(0.1396)	(0.1155)
First Year in Program	-0.0289	-0.0212
	(0.0809)	(0.0812)

Note: Each pair of cells represents the results of a separate regression of the treatment effect on the on an indicator for not being in the sample defined by the panel header. Column headers indicate the model components. Row headers indicate the baseline covariate with which the treatment indicators are interacted. N = 794. Models include randomization site fixed effects. Standard errors are clustered at the randomization site level. Source data in Panel A are parent surveys fielded in August of 2015. Source data in Panel B are teacher surveys fielded in May of 2015. Source data in Panel C are San Francisco Unified School District test files of the fall administration of the Fountas and Pinnell Benchmark Assessment System. Male and age in years are from district administrative Files. +indicates p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table A4
Heterogeneity of academic outcomes by year in experiment

	(1) General Text Treatment	(2) Personalized Text Treatment
Panel A: First Year in Program (N = 244)		
Reading level (ordinal logit)	-0.0381	0.5932*
	(0.2838)	(0.2524)
Reading level (standardized point scale)	-0.0494	0.2260+
	(0.1180)	(0.1271)
Exceeds Expectations	-0.0351	0.1575**
	(0.0540)	(0.0485)
Meets or Exceeds Expectations	0.0191	0.1087+
	(0.0602)	(0.0536)
Approaches, Meets, or Exceeds Expectations	-0.0012	0.0029
	(0.0534)	(0.0587)
Panel B: Second Year in Program (N = 344)		
Reading level (ordinal logit)	-0.0289	0.4324**
	(0.2614)	(0.1620)
Reading level (standardized point scale)	0.0223	0.1661
	(0.1027)	(0.1021)
Exceeds Expectations	-0.0017	0.0928*
	(0.0649)	(0.0432)
Meets or Exceeds Expectations	-0.0112	0.0868*
	(0.0497)	(0.0371)
Approaches, Meets, or Exceeds Expectations	0.0564	0.0127
	(0.0722)	(0.0519)

Note: Each pair of cells represents the results of a separate regression of the treatment effect on the relevant academic outcome. Column headers indicate the model components. The reference category is the control group. Row headers indicate the academic outcome. Panel headers indicate the quartile subsample. All models include randomization site fixed effects, controls for texting language, factors of baseline survey responses, and administrative covariates. Covariates are detailed in Table 3. Source data are district test files of the Fountas and Pinnell Benchmark Assessment System in fall of first grade. Standard errors are clustered at the randomization site level. +indicates $p<0.10$, *$p<0.05$, **$p<0.01$, ***$p<0.001$

Table A5
Descriptive Statistics for academic sample by quartile of baseline skills distribution (first year of experiment only)

	Quartile 1 (N = 59)		Quartile 2 (N = 61)		Quartile 3 (N = 62)		Quartile 4 (N = 61)	
Variable	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Panel A: Children								
Male	0.559	0.501	0.443	0.501	0.516	0.504	0.459	0.502
Hispanic	0.712	0.457	0.344	0.479	0.323	0.471	0.082	0.277
Asian	0.051	0.222	0.361	0.484	0.306	0.465	0.344	0.479
Decline To State Ethnicity	0.051	0.222	0.082	0.277	0.129	0.338	0.131	0.340
White	0.017	0.130	0.016	0.128	0.016	0.127	0.164	0.373
Other	0.169	0.378	0.197	0.401	0.226	0.422	0.279	0.452
Age in Years	5.520	0.334	5.469	0.281	5.456	0.280	5.554	0.288
Enrolled in Transitional Kindergarten	0.034	0.183	0.000	0.000	0.000	0.000	0.000	0.000
Parent rating of letter knowledge	2.356	0.689	3.070	0.750	3.746	0.472	3.754	0.537
Parent rating of letter sounds	2.881	1.068	3.164	1.067	3.645	0.907	4.090	0.990
Parent rating of rhyming	2.444	1.151	3.008	1.178	3.435	1.081	3.870	1.177
Panel B: Parents								
Has less than a bachelor's degree	0.746	0.439	0.672	0.473	0.677	0.471	0.377	0.489
Received Texts in English	0.322	0.471	0.590	0.496	0.516	0.504	0.721	0.452
Received Texts in Spanish	0.644	0.483	0.213	0.413	0.258	0.441	0.082	0.277
Received Texts in Chinese	0.034	0.183	0.197	0.401	0.226	0.422	0.197	0.401
Age in Years	32.568	6.407	34.538	7.530	35.089	7.166	36.142	6.261
How many times per week read to child	2.728	0.925	2.803	0.945	3.112	0.812	3.180	0.847
How many times per week told stories to child	2.554	0.751	2.684	0.904	2.694	0.801	2.959	0.848
How many times per week sang to child	2.610	0.905	2.799	0.891	2.774	0.857	2.692	0.812
How many times per week does child ask to be read to	2.745	0.863	2.836	0.986	3.048	0.876	3.212	0.824

Note: Parents rated the letter knowledge of their child in one of four categories: 1=The child knows no letters, 2=Some, 3=Most, 4=All. Parents rated how well their child can produce letter sounds and rhyme in one of five categories: 1=Not at all, 2=Not very well, 3=Somewhat well, 4=Well, 5=Very Well. Answer options for weekly parental activities and how often the child asks to be read to include: 1=Not at all, 2=Once or twice per week, 3=Three to six times, 4=Every day. Missing values set at the sample average. Baseline survey questions were given in September 2014. All child demographics come from San Francisco Unified School District administration files.