Using a Factorial Design to Maximize the Effectiveness of a Parental Text Messaging Intervention

Catherine Armstrong Asher
University of Michigan

Ethan Scherer
James S. Kim
Harvard University

Preprint dated: October 2021

Abstract**:** Parental text messaging interventions are growing in popularity to encourage at-home reading, school-attendance, and other educational behaviors. These interventions, which often combine multiple components, frequently demonstrate varying amounts of effectiveness, and researchers often cannot determine how individual components work alone or in combination with one another. Using a 2x2x3 factorial experiment, we investigate the effects of individual and interacted components from three behavioral levers to support summer reading: providing updated, personalized information; emphasizing different reading views; and goal setting. We find that the personalized information condition scored on average 0.03 SD higher on fall reading assessments. Texting effects on test scores were enhanced by messages that emphasized reading being useful for both entertainment and building skills compared to skill building alone or entertainment alone. These results continue to build our understanding that while text message can be an effective tool for parent engagement, the specific content of the message can lead to meaningful differences in the magnitude of the effects.

Keywords: literacy, parents and families, elementary schools, economics of education, experimental design

Acknowledgments: The authors are thankful for feedback and advice from Nicole Pashley, Carly Robinson, Luke Miratrix, Monnica Chan, Eddie Kim, Thomas Kelley-Kemple, participants in the READS Lab, CARES Lab, and conference participants at AEFP. This work is supported by the Chan Zuckerberg Initiative and the Institute of Education Sciences, U.S. Department of Education through grant R305B150010. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

Citation: Asher, C. A., Sherer, E., Kim, J.S. (2022). Using a factorial design to maximize the effectiveness of a parental text messaging intervention. Journal of Research on Educational Effectiveness, 15(3), 532-557. https://doi.org/10.1080/19345747.2021.2009073

Background

Across the social sciences, text messaging and other behavioral science interventions have experienced a rapid growth in popularity, with the promise of low-cost effects at changing individual behaviors. In education, both policymakers and researchers hope to use insights from behavioral sciences to enhance parental engagement and improve academic behaviors such as attendance, as well as academic skills like reading and math. As some efforts to scale and replicate the promise of nudging interventions have found null results, new questions have emerged about how to optimally design these behavioral interventions (Oreopoulos, 2020)**.** One promising subset of these programs provides parents with younger children regular and actionable text messages related to specific academic goals. Figure 1 provides a basic theory of change of how these behaviorally-informed messaging interventions could affect student outcomes. In this basic theory, each message provides the parent with an opportunity to change their action or communication with their child. We differentiate actions from communication since some messages focus on actionable tasks and others provide new perspectives and viewpoints that are not tied to a specific action. When parents observe the result associated with this decision, they evaluate whether the messaging intervention is helpful. Based upon the perceived helpfulness of the message parents may choose to continue responding (or not) to future messages. Over the course of the intervention, the accumulated changes in actions and communication between parents and students can affect student behavioral and reading outcomes. The regular and persistent nature of the messages means that no single message carries much weight in the decisions (or in the overall impact of the intervention) and allows parents to be "too busy" or to even disengage from the messages for a period until a more salient topic or message reengages them. This subtle difference could differentiate regular and

consistent messaging interventions from the specific behavioral science category of "nudges" popularized by Thaler and Sunstein (2009), where often single changes in decision-making frameworks can yield large down-stream effects.

This simple theory of change is particularly useful because it can be applied to many different types of messaging interventions aimed at changing parental behavior or communication. However, in education, learning to read is a fundamental focus of many educational efforts, and many messaging interventions target specific behavioral levers to change parents' decision-making around early literacy development. First, receiving a text message could refocus parents' attention to educational actions (Cortes et al., 2019; Doss et al., 2019; Hurwitz et al., 2015; Kim et al., 2019; Kraft and Monti-Nussbaum, 2017; Mayer et al. 2018; Smythe-Leistico & Page, 2018; York et al., 2018). Second, text messages can help parents overcome the challenge of short-term, immediate-cost behaviors (like reading to your child daily) whose benefits do not manifest until the future. Mayer et al. (2018) find setting goals and monitoring parents' progress towards these goals mitigated these self-control concerns. Third, a subset of the studies attempts to reduce the complexity, or cognitive load, of parenting by breaking down parental education activities into several discrete components (Hurwitz et al., 2015; Kraft and Monti-Nussbaum, 2017; Cabell et al., 2019; Doss et al., 2019; Kim et al., 2019; York et al., 2018). They provide a combination of literacy facts or resources, a specific activity or practice, and extension activities that parents can implement without additional preparation. Finally, informational messages attempt to facilitate parent monitoring of their children's academic performance. For example, it is difficult to track and monitor a child's cumulative absences, and parents often underestimate how much school their child has missed (Smythe-Leistico & Page, 2018). One method to address these concerns is providing parents with

information on their child's performance to correct these biased beliefs.[1]

Across these studies, most find positive effects on behavioral measures of parent or child reading, like the amount of reading, or parental involvement, between 0.15-1.0 standard deviations, (Doss et al., 2019; Hurwitz et al., 2015; Kraft & Monti-Nussbaum, 2017; Mayer et al., 2018; York et al., 2019) and Smythe-Leistico and Page (2018) find an 11 percentage point decrease in absenteeism. While many of them focus on pre-literacy skills often developed in pre-kindergarten (Cortes et al., 2019; Doss et al., 2019; Hurwitz et al., 2015; Mayer et al. 2018; York et al., 2018), others have focused on elementary-aged children aged 6-10 (Kim et al., 2019; Kraft & Monti-Nussbaum, 2017; Smythe-Leistico & Page, 2018). The varied effect sizes across both groups suggests more research is needed for the particular ages where consolidation of reading skills is a primary focus of learning. In one small-scale study finds that a messaging intervention paired with goal-setting and monitoring increased reading activities among families with young children by a full standard deviation (Mayer et al., 2018). The large magnitude of these effects represents a promising lever of change whose effects should be replicated at larger scale.

While it is clear from these studies that messaging interventions can change parental behavior, there is less clear evidence that they can influence more distal outcomes like student reading skills, as measured by test scores. While a few studies find positive effects of around 0.10-0.18 standard deviations (Doss et al., 2019; Kim et al., 2019; York et al., 2019) several find no significant differences or even mixed results for all children (Cortes et al., 2018; Cortes et al., 2019; Kraft & Monti-Nussbaum, 2017) and some do not examine test scores at all (Hurwitz et al., 2015; Mayer et al., 2018; Smythe-Leistico & Page, 2018). This variation could be related to different samples and context, but research increasingly suggests the importance of the specific

[1] Other studies have used informational messaging campaigns administered through a non-texting medium to reduce absenteeism among students (Rogers & Feller, 2018; Rogers et al., 2017; Robinson et al., 2018).

behavioral and content features of the message. Building on the work of York et al. (2018), Doss et al., (2019) find that an enhanced intervention with messages that differentiate activities according to students' baseline reading outperforms the intervention with undifferentiated s. While leveraging data to personalize messages has been shown to influence older children's academics (see Bergman, 2019 for a review) children's attendance (Rogers & Feller, 2018; Rogers et al., 2017; Robinson, et al., 2018; Smythe-Leistico & Page, 2018;), there has been no work, to our knowledge, on using information about student behaviors to personalize messages focused on the literacy skills of elementary students. The potential for personalized information to improve the efficacy of literacy-focused messaging interventions is one issue we address in this study.

Furthermore, while the existing research increasingly indicates the focus of the message content is important, interventions often combine message types or other components to maximize the intervention's effectiveness. In Doss et al. (2019), for example, the messages combine facts and information about reading with specific activities for parents to complete with their children in both arms of the intervention. Similarly, Mayer et al. (2018) combine their text messaging with regular check-in meetings with the program provider. Thus, while it is clear that the characteristics of the intervention matter, it is hard to assess whether individual components could work as well as a cocktail of messages or if the synergies between the components drive the effects. Unfortunately, typical randomized controlled trials, by design, are ill-suited to efficiently unpack the effectiveness of individual intervention components and their combination to answer the question of what works, for whom, and why.

While texting interventions can reduce specific behavioral barriers, less is known about how messaging interventions inform parent's domain-specific beliefs. For example, all the

previously mentioned literacy texting interventions focus the content of their message on building literacy skills, a framework aligned with the instrumental view of reading that emphasizes the value of skill-building for future success (Baker et al., 1997). However, literacy scholars document that while some parents emphasize an entertainment view, or reading for pleasure and enjoyment, as opposed to the instrumental view (Baker et al., 1997). Several small-sample studies find that parental beliefs differentially predict features of their children's reading experiences. Parents tend to promote types of reading activities for their children that align with their beliefs (Lynch et al., 2006; Sonnenschein et. al, 1996; Sonnenschein et. al, 1997). Additionally, parents who endorse the entertainment view have children who score higher on measures of reading enjoyment and motivation (Baker et al., 1997; Baker & Scher, 2002). Correlational research indicates that students with exposure to the entertainment view tend to be better readers than similar students from homes who emphasize the instrumental value (Baker et al., 2001; Sonnenschein et. al, 1997; Sonnenschein et al., 2000). Less is known, however, about whether these differences in views are causally related to student outcomes, and how underlying student motivation maybe be influenced by the introduction of different views. Causally investigating how literacy parental texting intervention effect could vary based upon promoting particular beliefs remains a critically unexplored area of literacy parental texting research.

The prior literature raises important unanswered questions for parental text messaging interventions about the specific behavioral and content features that make them most effective. Past work has suggested the importance of both data-driven personalized message content and goal setting, but no work has assessed these levers at scale individually or in combination for younger children. Furthermore, while the messaging to parents about reading beliefs has been pushed as a promising area of intervention, it has never been explored causally and never been

combined with other behavioral interventions. Our research seeks to understand what types of parental text messages are most effective, by investigating two primary research questions:

1. What components and combination of components of parental text messages affect parent behavior?
2. When do the effects on parental behavior transfer to student reading outcomes?

We address these questions by implementing a parental text message intervention to understand the effects of individual message components separately and in combination on family reading behaviors, student reading comprehension, and child reading behaviors. Specifically, we examine the effectiveness of three distinct components: personalizing messages with up-to-date information; the message's framing of reading as building skills, for entertainment or both; and setting reading goals at the beginning of the summer. We test these components of the text messaging campaign using a 2x2x3 factorial experiment with approximately 5000 elementary school students in a single school district in the southeast United States. A factorial design is particularly well-suited to study an intervention with multiple distinct components that could be individually included or excluded, because it allows for the estimation of each component's individual effects as well as interaction effects of multiple components (Collins et al., 2014; Somers, Collins, & Maier, 2014).

A unique feature of our intervention is our ability to both directly and indirectly measure parental reading behaviors in response to our text messages. While we follow prior studies and survey a subsample of parents, all participants in our study also had access to an educational reading app. Over sixty percent of our text messages encourage parents to have their child use the reading app. Login passcodes were only provided to the parents; thus, we can observe parental engagement when the child initially logs into the summer app. While students could

subsequently log themselves in, summer app usage data is a proxy for parental engagement since messages encouraged app use. We structure the remainder of the paper in the following way: first, provide more information about our intervention, data, measures, and methods used in our analysis; next, we present our main results and an investigation of potential mechanisms; we conclude with a discussion of our findings and their implications for texting interventions for parents.

The Intervention

Context and Study Eligibility

Students were recruited to participate in this study through their participation in a large, multi-school randomized control trial (RCT) of a curricular intervention. All first and second grade students in the participating schools were invited to participate in both the curriculum and texting studies through an active consent process, and families could choose not to receive messages. For the curriculum study, either first or second grade students at each school were assigned to receive the MORE (Model of Reading Engagement) intervention curriculum, with the other grade serving as a control group who received the business-as-usual curriculum. This intervention consisted of a series of science- and social studies-themed lessons in the spring semester and the students' choices of 10 hardcopy books related to the same science and social studies topics presented in the classroom lessons.

To be eligible for our text messaging study, the research team had to identify a valid cell phone number for a parent or guardian of each actively consented student (see the consort diagram in the Online Supplementary Materials, Figure S.1). Phone numbers were either provided directly on the consent form or by the school district administrative records. Once a cell phone number was validated, the families were enrolled in the text messaging study.

Participating families were then randomized into conditions that received different versions of the texting intervention, including a small group that received no text messages.

At the end of the school year, all students in both the curriculum study and texting study received access to an educational reading app called MORE@Home, which contained six digital books as well as a series of reading-related activities matched to each book and broadly leveled according to the child's end-of-year reading ability. For those students in the MORE curriculum arm of the curriculum study, their MORE@Home accounts also provided access to leveled reading activities for each of the 10 books they had selected. On the whole, use of the app was relatively low across the sample. Among our pure control group of families, only 16% of student accounts were ever activated by their parents.

Text messaging intervention

The purpose of the text messaging intervention was to increase parental engagement with their children in summer reading activities and improve student learning. In line with recent research on the effect of frequency and timing of text message interventions (Cortes et al., 2018; Cortes et al., 2019), families received text messages twice weekly over 9 weeks of summer vacation, with one message occurring earlier in the week and one message closer to the weekend. Messages were sent in either English or Spanish, based on the student's home language in district administrative records.[2] Messages were all sent through the Twilio messaging platform which was accessed through our sample database. This approach had several advantages: 1) messages were linked to user profiles containing other research information from the curriculum study and educational app database; 2) families could easily opt-out of messages by replying with "STOP", preserving consent; and 3) parental responses to text messages were logged on the

[2] Besides English and Spanish, no other home language represented more than (1%) of students. Families whose home language was neither English nor Spanish received English messages due to resource constraints.

user profiles. Some text messages were designed specifically to promote usage of the educational reading app, while others encouraged a wider variety of reading activities. Each message contained a single topic, which generally covered one of the three larger themes: 1) reminders to engage in summer reading activities; 2) providing information about summer reading resources (including the educational app); and 3) monitoring progress throughout the summer.

Differentiating messages

To explore how text messaging features differentially influence parental and student engagement with a summer reading intervention, we differentiated the specific wording of each text message topic according to three separate factors that have shown promise in prior research: updated personalized information to correct parent misbeliefs (Smythe-Leistico & Page, 2018), goal-setting to reinforce immediate action (Mayer et al., 2018), and framing different views of reading to promote specific parental beliefs (Baker et al., 2001; Sonnenschein et al., 2000).

In the personalized information factor, some families received text messages that include student-specific information within the message. Some examples of the information that could be included were: the specific books the student had access to in the app, whether or not a student had logged into the educational app yet, or which books' activities the students had accessed on the app. Because the text messaging was integrated with the app's backend database, each student's information was updated continuously, ensuring that the messages reflected the students' most recent status. Families not in the personalization condition received more generic messages, but they still referred to individual students by name.

For the goal-setting factor, some families were invited to set a summer reading goal at the beginning of the intervention, with later messages periodically checking-in on their progress towards that goal. We designed the goal setting to be a light-touch, low-cost, scalable version of

other effective goal-setting studies (e.g., Mayer et al., 2018; Oreopoulos et al., 2020), with parents being asked not only to identify a goal but to make a plan for reaching it in the face of obstacles (Oettingen & Reininger, 2016). However, without an in-person goal setting session or subsequent individualized follow-up, most families (more than 95% of the goal-setting condition) failed to complete the goal-setting exercise. For the families who did set a reading goal, check-in messages would explicitly refer to their individual goal. Other messages referred to "your summer reading goal." Often in education and the social services, policymakers cannot force individuals to take advantage of specific opportunities or services. Thus, even with the low rates of follow-through, providing families with the opportunity to set and monitor a goal represents a realistic, or ecologically valid, intervention whose results will be relevant for districts considering whether to implement a similar program with few monitoring resources.

Finally, the view-of-reading value's condition created three distinct groups who received differently framed messages over the course of the campaign. The instrumental-only-view condition emphasized reading in order to develop specific skills important for future success. The entertainment-only-view condition emphasized reading as an enjoyable and fun activity. The combination-view received a balanced combination of entertainment- and instrumental-framed messages over the course of the summer. Importantly, however, this combination-view received the same total number of messages as the entertainment-only and instrumental-only conditions.

Not all messages included components based on the levels of all 3 factors. A single message could meaningfully differentiate between conditions for just one factor or across any combination of two or three factors. In our text messages, 40% were relevant to families' goal-setting condition, 47% differed by personalization, and 85% were framed for a specific view of reading. To ensure that the intended conditions were salient, we recruited colleagues to review

example text message versions and provide feedback on whether the messages clearly contained information aligned to specific conditions. For most piloted messages (67%), raters were able to correctly identify the differentiation. In cases where at least two colleagues' perspectives differed from our intention, we revised the message versions to increase or decrease the salience of a specific condition. A comprehensive list of message themes as well as example message variations can be found in the Online Supplementary Materials (Tables S.1 and S.2).

Methods

Sample

This study includes 5,172 rising second and third grade students, from 4,993 families, who attend thirty elementary schools in a large school district in the southeast. Demographic characteristics are presented in the first column of Table 1. Close to forty percent of the students are African American and an additional thirty percent are Hispanic. Approximately twenty percent are white, and approximately ten percent are Asian. Almost a quarter of students were receiving English-learner services. Our sample contains socioeconomic diversity but contains a larger proportion of students in low-SES neighborhoods relative to the district as a whole.

Research Design

We use a factorial experiment to compare the differential effectiveness of text messaging components. Traditional randomized controlled trials are only able to compare two treatment arms at a time, so investigating three potential mechanisms would require multiple experiments, or a multi-arm RCT with a prohibitively large sample size. A factorial design, however, is particularly well-suited to study an intervention with multiple distinct components. In a factorial design, each intervention component is treated as its own factor, with different levels representing the treatment assignment. Each unit is randomized to a level for each factor

independently. This design has the benefit of allowing the researcher to test the main effect of each intervention components on its own, as well as interactions of intervention components (Collins et al., 2014; Somers et al., 2014). It is thus an appropriate design to address how the multiple levers targeted in texting messaging interventions contribute to an intervention. We use a full-factorial design, in which every factor is fully interacted with the other factors. The goal-setting and personalized information factors each have two levels (on, off), and the view-of-reading factor has three levels (instrumental view only, entertainment view only, both views presented), resulting in 12 different treatment combinations. Additionally, we assigned a small portion of the sample to a pure control condition, not receiving any text messages. Separating out this pure control provides a business-as-usual condition to use as a benchmark for the magnitude of the factorial differences, but our research questions focus exclusively on the relative effectiveness of the different text message components as opposed to the effects of text messaging compared to no messaging. Table S.3 in the Online Supplementary Materials shows the full set of treatment groups within our sample.

To account for the presence of siblings in the sample, which could result in spillover and confusion for parents receiving two types of messages, random assignment to conditions occurred at the family-level. To improve precision and reduce the minimum detectable effect size, the sample was blocked at the school-by-grade level, the unit of treatment from the larger RCT, because average student reading levels and implementation fidelity of the larger intervention vary across schools. Within these blocks, the 4,993 families (5,172 children) with valid cell phone numbers were assigned to one of the 13 conditions.

A series of balance checks are presented in the remaining columns of Table 1, comparing our sample across each factor of the intervention. Overall, we find few differences between

experimental groups on baseline demographic characteristics or academic performance, and none that are statistically significant after applying a multiple hypothesis correction to account for the number of student characteristics compared (Benjamini & Hochberg, 1995). The similarities between groups reflects a successful randomization process.[3]

Data Sources

Baseline data from district administrative records of the pre-intervention year (2018-2019) is available for our entire study sample. These measures include enrollment information, student demographics, and reading and math test scores. The school district where this intervention took place does not collect student-level measures of socio-economic status; however, student neighborhoods, as determined by their census block-group, are categorized as being low-, middle-, or high-SES communities. Student-level outcome data from Fall 2019 is also provided by the district. We use the Measure of Academic Progress RIT score in literacy as (MAP, Northwest Evaluation Association, 2011) a primary outcome, which measures foundational literacy skills in the domains of literature and informational comprehension, vocabulary, phonics and decoding, and basic writing conventions. A second academic outcome, available for the rising third-grade cohort, is the Beginning of Grade (BOG) assessment, a statewide measure designed to measure student's preparedness for the third grade End of Grade (EOG) accountability assessment. The BOG assessment addresses multiple domains, including reading for literature, reading for informational text, and language, and students are required not only to "recall information, but also apply concepts and skills, make decisions, and explain or justify their thinking," (NC DPI, 2020, p.2). Both assessment scores are standardized within grade to provide outcomes in standard-deviation units.

[3] Though not displayed in Table 1, we also conducted balance checks between our message-receiving conditions and the pure control. We found no significant differences on any pre-treatment characteristic.

We assess the effect of the intervention on parents' behaviors and beliefs with two sources. First, we track family use of the MORE@Home app. Because 60% of the text messages parents received were related to the educational app, usage statistics reflect whether parents changed their own and their children's behavior in response to messages. While students could use the app independently, parents received the initial passcode with their student's unique login and needed to share that with for child to initially access the app. Additionally, most students in our sample did not have their own device and needed to use a guardian's device to access the app. For these reasons, we characterize the use of the app as an indicator of parental behavior. Specifically, we are interested in whether parents logged their students into the app, the total number of books they accessed on the app, and the total number of minutes they spent engaged with the app. For families who never logged into the app, total books and total minutes were recorded as zeros. We also collected self-reported outcomes from a subsample of parents who were randomly invited to complete a parent survey. The survey, which replicated a parent survey in Kim et al. (2019), included questions about their summer reading activities outside of the educational app and their perceptions of the text messages. A composite index using four distinct items had a Cronbach's alpha of 0.71 in our sample.

To explore student-level mechanisms and provide more information about students' summer reading experiences, we collected two waves of student surveys as well as qualitative measures of the students' educational app use. In the spring prior to the summer texting intervention, we measured student motivation using the Me and My Reading Profile (MMRP, Marinak et al., 2015). In the fall, we administered a student survey that included the MMRP and additional questions about students reading behaviors over the summer. The fall survey was administered in the thirty study schools and thus does not cover either the students who moved

out of the district or to a non-study school within the district. For students whose parents logged them into the app, we measured the percentage of activities that they completed correctly. We also asked periodically whether they enjoyed the app, felt like a good reader, and found the app activities challenging. Because our sample is slightly different for each set of outcomes, we also test for differential attrition rates based on the different factors of our treatment. We find no significant differences in retention rates by condition for test scores, parent and student surveys, and objective app metrics. A consort diagram showing our randomized sample and the sample included in each analysis as well as a table of the results from the attrition analysis can be found in the Online Supplementary Materials (Figure S.1 and Table S.4).

Empirical strategy

We analyze the twelve conditions in our factorial experiment in an intent-to-treat analysis using standard regression techniques. In a recent paper reviewing recent factorial studies in economics, Muralidharan et al. (2020) find that leaving out treatment interaction terms yield incorrect inferences if the interaction effects are non-zero or if model-selection is determined after looking at the magnitude and significance of these interaction effects. To avoid both of these concerns, and to allow for concurrent interpretation of main effects and interaction effects of our different factors, we use the following model:

$$\begin{aligned} Y_{ij} = {} & \beta_1 Pers_{ij} + \beta_2 EntView_{ij} + \beta_3 BothView_{ij} + \beta_4 Goals_{ij} + \beta_5 PersxEntView_{ij} \\ & + \beta_6 PersxBothView_{ij} + \beta_7 PersxGoals_{ij} + \beta_8 GoalsxEntView_{ij} \\ & + \beta_9 GoalsxBothView_{ij} + \beta_{10} PersxEntViewxGoals_{ij} \\ & + \beta_{11} PersxBothViewxGoals_{ij} + \Gamma X_{ij} + \phi + \varepsilon_{ij} \end{aligned}$$

where Y_{ij} represents the outcome for individual i in randomization block j. The model also includes a vector of covariates X_{ij}, including student demographics and pre-test math and

reading scores, as well as a set of fixed effects, ϕ, representing the randomization blocks. Standard errors are clustered at the school-grade level as recommended by Athey & Imbens (2017), to account for the unit of blocking and the correlation of residuals within those blocks.

The main effects (captured in β_1 through β_4) provide information about the average effect of each text messaging factor, and the two-way interactions (captured in β_5 through β_9) tell us whether the effects of one factor depend on the levels of the other factors. The three-level view-of-reading factor has been separated into two variables, EntView (for entertainment only) and BothView (for a combination of values), with the instrumental view-of-reading condition serving as the reference category for both of those variables. The three-way interaction terms (β_{10} and β_{11}) are included to allow us to estimate the main effects and two-way interactions concurrently; given their difficulty of interpretation, they are not parameters of interest for this paper.

To facilitate interpretation of effects, treatment variables are coded using effect coding with variables taking on values of -1 or 1 (Kugler et al., 2012), as shown in Table S3 in the Online Supplementary Materials. Because our sample is evenly divided across conditions, this parameterization allows both the main effects (the difference between levels of each factor or marginal effects) and the interaction effects (the additional effect of receiving a particular combination of factors) to estimated concurrently (Hardy, 1993). However, because level differences require the treatment indicators moving from -1 to 1 instead of from 0 to 1 as in the standard dummy coding, we must multiply all coefficients (and their standard errors) by 2. Thus, we can interpret the main effect β parameters as follows:

- The average effect of receiving personalized information (vs. not) is $2\beta_1$.
- The average effect of framing messages with the entertainment views of reading (compared to instrumental-only view) is $2\beta_2$.

- The average effect of framing both views of reading (compared to instrumental-only view) is $2\beta_3$.
- The average effect of setting goals (vs. not, as if in a traditional RCT) is $2\beta_4$.

Interaction terms are interpreted as normal once they have been scaled up: the additional effect of one factor in the presence of another factor. In all tables in this paper, we have already adjusted the point estimates and standard errors for ease of interpretation. Because we present multiple outcomes in each domain, we also test our confirmatory results to the sensitivity of false discoveries, using the Benjamini-Hochberg procedure with a false discovery rate (FDR) set to 0.05 (Benjamini & Hochberg, 1995) by outcome domain.

Results

Effects on parental behaviors and beliefs

We first consider how the different types of text messages affect parental behaviors and beliefs in Table 2. Panel A presents the main effects of each component, and Panel B presents the interaction effects, but for each column, both panels come from the same fitted model. Our primary measures of behavioral change in parents are captured in their use of the educational reading app with their children. For both the probability of ever logging into the app, the total minutes spent on the app and the number of books completed, the personalized messages were significantly more effective than non-personalized messages. Families receiving personalized messages were three percentage points more likely to use the app (ES=0.08, $p<0.05$), spent about an extra 1.6 minutes using the app (ES = 0.11, $p<0.01$) and completed an additional 0.7 books worth of activities (ES=0.12, $p<0.01$). These effects remain statistically significant after multiple-hypothesis corrections (Benjamini-Hochberg, 1995). We find no evidence of main effects on family app use from providing parents with the opportunity to set a summer reading

goal or from changing the view of reading that was emphasized.

In Panel B, we see that the goal-setting and view-of-reading have significant interaction effects. The effectiveness of a combination of values increases app usage when combined with goal setting, even though the individual components were not significant on their own. The combined effect of goal-setting and a combination of view-of-reading was significantly more than either of the individual effects of app usage: 4 percentage points higher login rate, 1.8 additional minutes on the app, and an addition 0.7 books completed. These effects are similar in magnitude to the main effects of personalization described above but need to be considered in the context of the non-significant main effects of the goal-setting and view-of-reading factors. Thus, goal-setting alone does not seem to have an effect on directly observed app usage, but the combination of goal setting and changing the view of reading does affect these outcomes.

To facilitate interpretation of our effects, Figure 2 presents model-based predicted outcomes for our sample with confidence intervals, grouped first by personalization and the view of reading and aggregated across the goal-setting condition. The purple dashed lines show the average use of the educational app by students in the pure control condition. Sixteen percent of these families ever logged into the app. On average, they spent about 5 minutes using the app over the summer and completed just under 2 book's worth of activities. The personalized text results are in green circles and the generic texts are in blue squares, and the horizontal groupings represent the view-of-reading conditions. From these, we can more easily identify common trends that are not immediately apparent from the model output. This figure highlights that personalization improved app use for all three view-of-reading conditions, but that compared to our pure control condition, both generic and personalized messages were effective at increasing parental engagement with reading activities. These results are consistent with prior absenteeism

work using postcards that providing personalized information to parents debiases their belief when information is hard to obtain (Robinson et al., 2018; Rogers & Feller, 2018) and show that personalized text messages caused parents to login and use the reading app more.

We also consider self-reported behaviors and beliefs from the subsample of parents who responded to our parent survey. These results are presented in the right-hand panel of Table 2. Because the survey sample is relatively small, we only examine the main effects of each factor, and we find limited evidence of effects. Among this group, relative to those receiving instrumentally framed messages, receiving only entertainment-framed messages decreased the frequency of reading by 0.27 standard deviations ($p<0.10$). Providing a goal-setting opportunity increased the likelihood that parents found the text messages useful by 32 percentage points ($p<0.10$), as did receiving messages framed around both views of reading compared to only the instrumental view (by 36 percentage points, though not statistically significant). We speculate that the divergence of app and self-reported results could be because the content of these messages change parental activities beyond the limited behaviors that we can observe in the app.

Effects on student reading performance

To understand whether these changes in parental behaviors and beliefs translate into effects for their students, we present effects on reading scores in Table 3, following the same panel structure to consider both main effects and interaction effects simultaneously.

We see evidence of transfer from the behavioral effects of personalized information to student test score outcomes. Personalized text messages significantly improve Fall MAP scores by 0.03 standard deviations ($p<0.05$). While not significant, the point estimate for the effect of personalization on Beginning of Grade (BOG) scores is also positive, though slightly smaller than the effect on MAP (ES = 0.02, $p<0.10$). The MAP effect remains marginally significant

after correcting for multiple hypotheses. We find no significant main effects for changing goal-setting or view-of-reading components on student test scores. We interpret the positive effects of personalization as transfer of the increased family reading activities, because the app was aligned the MAP assessment. The BOG effects represent farther transfer, which may be why they are smaller and less precise.[4]

Panel B shows the interaction effects when we consider whether the observed main effects of each factor depend on the levels of the other components. Though the point estimates are only significant for Fall BOG scores, we see a similar pattern across both outcomes: personalization with the entertainment-only view is slightly less effective than personalization with instrumental only view (for the BOG, ES = -0.077, $p<0.05$), but personalization with both reading views presented is more effective than personalization with instrumental only values emphasized (for the BOG, ES = 0.104, $p<0.001$). Taken together with the positive interaction effects of the combined view and goal setting on parental reading behaviors, these effects on student test scores are further evidence that the view-of-reading framing of messages has the potential to enhance or detract from the effects of other factors.

Figure 3 shows these results visually and replicates the layout of Figure 2. From the figure it is easy to see that personalization outperforms no personalization because the green circles are almost always above the blue squares across MAP and BOG. While we saw from the output that personalization improves test scores overall, Figure 3 shows that this effect is two to three times larger for students receiving a combination of reading values than for those receiving a single type of reading value, reflecting the significant interaction effects. For the MAP, this extra gap comes mostly from improved outcomes in the personalization x combination-of-views

[4] The differences in effects between the two outcomes are not due to differences in sample; in sensitivity analyses we find similar effects on the MAP when we limit the sample to rising third graders who also took the BOG.

group, whereas for the BOG, the amplification also results from the no-personalization x combination-of-views group underperforming relative to the other no-personalization conditions. In this figure, the purple dashed line representing average outcomes for our pure control group highlights that the specific content and framing of the message matters. While the most effective types of text messages are helpful for student reading, less-effective messages may be worse than no messages at all. The most consistent way to ensure that a messaging campaign is helpful is to use it to provide student-specific information to parents.

Potential mechanisms

Next, we investigate potential mechanisms for the transfer of effects from parents to students. We explore why the effects of view-of-reading content that parents received in text messages only appear to amplify the effects of message personalization for student reading. Table 4 explores the effects of our messaging components on alternate student outcomes. Students whose parents received personalized messages reported reading approximately 2.5 percentage points more of their available books on the MORE@Home reading app.[5] There are no significant main effects for either goal setting or view-of-reading on self-reported reading. Looking at the interaction effects for the proportion of books read, we continue to see that emphasizing the entertainment value tended to mitigate, or counteract, other effects, particularly when combined with goal setting. Students in both the goal-setting and the entertainment view-of-reading conditions read 3.1 percentage points fewer of the available books than would have been predicted had they received goal-setting and the entertainment-views on their own.

We see a different pattern when we consider the differential effects on students' reading

[5] As described previously, all students received access to 6 electronic books and activities via the MORE@Home app, but half also received a set of 10 hardcopy books and additional app activities. Thus, for one half of the sample, the percent of available books read is a proportion of the 16 total books they had access to via the program.

motivation as measured by the MMRP. There are very few significant main effects, though students receiving a combination of reading values emphasized had fall reading motivations that were lower (ES = -0.08) than students who only received instrumental messages. This negative effect of the combination framing is amplified among students who were also in the goal-setting condition (ES = -0.13). On the other hand, the combination of goal-setting and the entertainment value increased student reading motivation (ES = 0.10). Although these patterns are not precisely estimated, the results are particularly interesting because they point in the opposite direction of our main test score results. Students who received a combination of reading values in their messages in addition to either goal-setting or personalization used the educational app more and scored higher on reading assessments. But these same students reported lower reading motivations when they returned to school in the fall. These results suggest that despite the unchanged family behaviors in the app, the framing messages did change the student reading experience outside of the app, possibly through parent communication to their children.

We continue to explore the motivation hypothesis by exploring what students reported about their experiences with the MORE@Home reading app. We know that the app users are a relatively small and unrepresentative subset of our total sample – because personalized messages increased app usage, our app-user sample is overrepresented by families from that condition. Overall, we find no evidence of main effects of our message components on app user's experience. Students generally enjoyed the activities and felt like good readers about the same amount across all conditions. However, the interaction effects in Panel B align with the self-reported reading motivation results. Students in entertainment value and personalization conditions felt like much better readers (ES = 0.19) than students who were in instrumental value and personalization condition. On the other hand, students in both the combination of values and

personalization conditions reported feeling like worse readers than students who were in one or the other (ES = -0.15, p>0.10). This may indicate that students who were encouraged to use the reading app for fun felt slightly better about themselves afterwards, whereas students who received targeted "mixed messages" about enjoying and learning from the app tended to feel slightly worse. While we know that the app users are a relatively small and unrepresentative subset of our total, these exploratory findings reinforce the idea that parental communications to children may have been influenced by message content (the view of reading) in a way that shaped students experience beyond just increasing app usage.

Discussion

In this study, we demonstrate how a factorial design applied to a parent literacy texting intervention can improve both parental engagement with summer reading resources and student academic outcomes. The main contribution of our work is that it clearly demonstrates that variation in the content and tone of messages can lead to important differences in effects. Moreover, our use of a factorial design allows us to estimate not only the impact of each individual component, but also how those components' effects interact.

We find that data-driven personalized messages improve parental use of an educational reading app across a variety of measures and transfer to positive effects on student reading performance in the fall. Our findings that more personalized messages outperform generic messages on reading comprehension outcomes is consistent with prior work on literacy reminders (Doss et al., 2019), but also differs from their research in important ways. Doss et al. (2019) varied the parents' recommended reading activities based upon baseline performance of individual students. Our study relied on the educational app to provided leveled activities for student learning across all study conditions. This may be one reason that we find smaller

differential effects than Doss et al. (2019). Another difference that could have led to smaller effect sizes than Doss et al. (2019) is the shorter duration of our texting campaign. Our texting intervention was only 9-weeks during the summer compared with a 10-month period encompassing the school year and the summer. Parents are saturated with texting campaigns and thus, it could be that with deeper personalization over a sustained time-period, allows one to establish themselves as a trusted source of information and rise above the noise.

The nature of our personalization was also different than in Doss et al. (2019). In our messages, the personalization factor adds relevant and timely information to parents about their student's book selection for the summer and use of the educational app. We hypothesize these information effects could have emerged through two different mechanisms. First, our personalized information texts caused parents to log-on to the educational app more and our point estimates, while not significant, suggest that parents engaged slightly more frequently in reading activities with their children outside the app as well. The real-time information about log-ins and books could be correcting parent's beliefs about how much time they spend on reading activities with their child over the summer. Prior work demonstrates that parents are not always informed about their children's academic behaviors. For example, they tend to underestimate school absences (Robinson et al., 2018; Rogers & Feller, 2018; Smythe-Leistico & Page, 2018). Providing up-to-date information about the number of days a child missed during the school year helps correct these parents' misbeliefs about student absenteeism. In our context, this suggests that parents may be overestimating how much time they are spending with their child on reading over the summer and the information helps correct these biases.

Another helpful component of our interventions could have been that the content of our messages was not focused exclusively on reading activities, but rather included messages related

to reading resources, reminders to engage in reading activities, checking-in on progress and in some cases framed reading as both skill building and entertainment. Parents might hit barriers as they try to implement different strategies and thus giving parents an opportunity to gain confidence in some areas, could help build self-efficacy. Recent work suggests that varying content by domain could be beneficial in sustaining parent's attention on longer campaigns and increasing their self-efficacy in parenting (Doss et al., 2018; Doss, Fricke, and Loeb, 2020).

Our experiment is also the first to rigorously evaluate hypotheses emphasizing particular reading views. Home literacy theory suggests that parents' emphasis of an entertainment-view of reading could lead to improved student motivation and literacy outcomes (Baker et al., 2001). We find some evidence that an entertainment framing, combined with either goal-setting or personalization, increased student reading motivation and the positive feelings about reading. However, we do not see broad evidence of transfer to self-reported reading motivation of the entertainment-framing alone and find no average effects on student test scores. Thus, while we do find that presenting a combination of the entertainment and instrumental views amplifies the positive effects of personalization, and only presenting the entertainment view detracts from the effects of personalization, we do not have sufficient reason to believe that these effects operate through, or are mediated by, student motivation, as would be suggested by prior research on home literacy (Baker et al., 2001). In summary, students in the personalization and entertainment condition saw statistically significant and positive effects on whether they felt like a good reader, as well as positive effects on a survey of reading motivation, with the opposite being true for when both the entertainment and instrumental views-of-reading were present. While entertainment-themed messages increased student motivation, this did not translate into the types of behaviors associated with improving reading skills. All in all, this study does not provide

support for the home literacy theory of how to improve student reading performance.

Another paradoxical finding relating to home literacy theory is that while we see some evidence of the benefit of the entertainment view, we also find that a combination of the entertainment and instrumental views lowers reading motivation. One potential reason for this disconnect between the motivation and test score outcome relates to the contradiction between the messaging and actual experience of the students. It could be that adding instrumental messages encouraged parents to practice skill-building with their children, which can be less enjoyable, but would improve their reading scores. Students might experience a disconnect between their expectation of pure enjoyment and the effort required to use the app and practice challenging reading activities. While we hesitate to extrapolate too much from these exploratory and inconsistent findings, it raises the questions for future research about the importance of aligning expectations and experiences in messaging interventions.

We were surprised by the limited effectiveness of the goal-setting component of our intervention, which tended to have point estimates close to zero across multiple outcomes. Mayer and colleagues (2018) found large positive effects on the use of an educational app in their comprehensive behavioral intervention (PACT) that leveraged goal setting, reminders, and social pressure. However, there are two key differences that may explain these discrepant findings. First, the goal-setting component of PACT was intensive, involving weekly face-to-face meetings with program staff, which was possible because they were only working with 80 students. Our goal-setting intervention was much lighter touch, because it had to be scaled up to reach almost 5,000 families. With so few parents electing to set goals through our online form, the goal-setting factor of our intervention should best be described as a goal-oriented framing of summer reading as opposed to the commitment device used in PACT. The lack of evidence in

support of this effectiveness suggests that the framing is not sufficient to change family behaviors, and the intervention must be sufficiently intensive to act as a commitment device. In addition, goal setting was just one of three major components of the PACT, along with the educational app and social incentives. We have seen from this study that our own positive effects were largely driven by information-sharing, which aligns more closely with the reminder and social-pressure components of the PACT intervention. It is possible that the large positive effects Mayer et al. find are primarily due to the other components of the intervention.

More broadly, the low take-up of goal-setting reflected a broader pattern of low family engagement with the educational app across all study conditions, and the challenges parents may confront during the summer months. This was most extreme for the pure control group, where only one-sixth of families ever used the app, but low take-up was also a concern across all of our treatment groups. While it is promising that even in this context we observed significant changes in behavior and small, but still significant improvements in student test scores, more engagement would have been better. In their summer messaging intervention, Kraft and Monti-Nussbaum (2017) found that common barriers to engaging in literacy activities among families included summer plans, family challenges, and work demands. Other research on summer literacy suggests that habit formation at the end of the school year may help students continue with specific reading behaviors over the summer (Kim et al., 2016; Kim et al., 2019). One possible way to improve engagement would be to embed use of the educational app within the classroom, or for teachers to help students and families develop specific habits before summer starts.

Policy Implications

The implications of our study are particularly relevant when considered alongside recent syntheses that highlight the difficulty of scaling successful behavioral interventions. For

example, recent work by DellaVigna & Linos (2020) examined 126 nudge randomized controlled trials in government "nudge units" covering over 23 million individuals. Across their sample, they find statistically significant, but considerably smaller, effects on nudging interventions when compared to the published literature, which they attribute to publication bias and power (DellaVigna & Linos, 2020). However, while the authors do separately assess the mode of communication and the underlying behavioral mechanism, they do not explore content differences, which our work shows could be a crucial feature. As we saw in Figure 3, we find that non-personalized messages perform worse for student test scores than sending no text at all, whereas the personalized messages are clearly better. In other successful messaging interventions to increase college enrollment, personalized messages to the students and from the institution staff are standard (see Bird et al., 2019 for a review of recent papers). Even though recent work by Bird et al. (2019) tests how these messaging campaigns scale for 800,000 students and find discouraging precisely estimated null effects, they also acknowledge that their ability to personalize the messages was constrained as the intervention scaled. Our personalized information relied on data from the app as well as knowledge of the books available to the children. Other forms of data, like learning management systems, have been leveraged in other studies (see Bergman, 2019 for a review). These patterns suggest that integrated data-systems could become increasingly important and valuable tools for changing behavior.

Additionally, in education there has been little causal research on whether changing how interventions are framed for participants could influence behavior and decision-making. For example, all the prior intervention literature on parental texting for younger children focuses exclusively on communicating with parents about building reading skills. Our work, suggesting that messages can broaden parents' choice sets by exposing them to additional views, could be

important for future research for both text messaging and educational messaging more broadly.

Relatedly, we show that not only does the content of the message matter, but also that it is insufficient to consider how elements of message content like personalization or message-framing operate independently. For our most distal outcome, BOG scores, neither personalization nor an emphasis of particular view-of-reading yields large, statistically significant effects on its own. However, combining personalization with a combination of reading views is significantly more effective than either one of them alone. We hypothesize that message framing seems particularly important for more distal outcomes because the nuanced differences highlighted in these messages could be particularly important for parental participation in reading behaviors unrelated to the app, such as reading together or talking about reading in the home. The factorial design, which allowed us to show that not all messages are equally effective, can be a valuable tool for other researchers who are hoping to identify the most promising features of a messaging intervention in a variety of domains.

Another key difference between some of the recent attempts to scale messages and successful early literacy interventions is the sustained and consistent contact with parents that we highlight in our theory of change. These sustained campaigns are structurally different that typical behavioral nudges, particularly in a family outreach context, because they offer more opportunities for engagement than single or short-term messaging campaigns. The drawback is that a sustained intervention must remain relevant to families. One way to sustain the relevance of a messaging campaign is to broaden the scope of the messages. York et al. (2019) find that their effects of the intervention were much stronger in the second year of implementation, when the messaged touched on a variety of topics, including literacy, mathematics, and socioemotional skills, rather than literacy alone. While they are unable to test the hypothesis directly, it seems

likely that the length of the intervention (eight months) benefited from this additional variety and would have otherwise felt repetitive. More recently, Doss et al. (2020) also find evidence of the importance of varying the domain of the content for young students. Similarly, in addition to the app-related resources and reminders, our most successful condition (a combination of reading views in addition to personalized information) included multiple reasons for and ways to engage in reading. Applying these lessons could help text messaging interventions move beyond "nudges" into more sustained assistance and support for parents.

Finally, our work continues to build on the existing literature that a behavioral messaging intervention can improve student outcomes at a much lower cost than other interventions with similar effect sizes. We estimated a back-of-the-envelope cost of this intervention of less than $4 per student, which included the cost to build the connection and integration between a student database and the texting software, the staffing associated with composing each text message, and responding to questions and messages received from families, as well as the cost of sending the text messages through Twilio's platform. Our personalized messages are conditional on other infrastructure to easily provide data, like the app in this study, or a learning management system in others. However, presumably, systems like these are installed for other district reasons and thus we exclude their cost. While our test score effect sizes are modest, the cost-effectiveness would compare favorably with the Tennessee STAR class size experiment (Schanzenbach, 2006), and other literacy interventions, including Reading Partners (Jacob et al., 2016), Project READS (Kim et al., 2016), and others (Hollands et al., 2013).

Future directions

Our work also offers several lessons for the design of future evaluation of parental texting interventions. First, our text messages were direct to parents, yet only a small proportion of the

families responded to our survey, which limited our ability to unpack some of the mechanisms that drove our results. Because factorial design allows researchers to explore heterogeneous effects, ensuring sufficient effort to collect this data source will be important for future work.

Similarly, we solicited our initial goal-setting survey via text message and received a poor response rate on the parental goals, despite being based in goal-setting theory about the importance of plans (Oettingen & Reininger, 2016). Perhaps an alternative in-between our work and Mayer et al., is to call parents for the survey. In other fields, like political science, phone-based interventions encouraging voters to make a plan to vote significantly increased turnout (Nickerson & Rogers, 2010; Rogers et al., 2015) and now is common practice in voter outreach.

Third, as noted in other work (York et al., 2019; Kraft and Monti-Nussbaum, 2017) there is a learning curve to sending these parental text messages at scale, thus we suggest that district and policymakers commit to these interventions in order to reap the benefits after the setup cost.

Finally, several recent papers have found smaller or even null effects when text messaging campaigns are scaled to tens or hundreds of thousands of students (Bird et al., 2019; DellaVigna & Linos, 2020), and our own study also struggled to generate parental buy-in when we scaled a goal-setting intervention in a light-touch medium. However, our personalized messages using student's book selection for the summer and app data caused cost-effective improvements in test score outcomes even though our sample was larger than some of the previous literacy-focused studies. While our sample was not on the same scale as Bird et al. (2019), the integration of our data and messaging systems would not face the same constraints of limited information that they found relying on state- and national-level data to inform their student messages. Thus, further investigation is needed whether scaling using these detailed district-level data to personalize messages can be effective.

References

Athey, S., & Imbens, G. W. (2017). Chapter 3—The Econometrics of Randomized Experiments. In A. V. Banerjee & E. Duflo (Eds.), *Handbook of Economic Field Experiments* (Vol. 1, pp. 73–140). North-Holland. https://doi.org/10.1016/bs.hefe.2016.10.003

Baker, L., Mackler, K., Sonnenschein, S., & Serpell, R. (2001). Parents' interactions with their first-grade children during storybook reading and relations with subsequent home reading activity and reading achievement. *Journal of School Psychology*, 39(5), 415–438. https://doi.org/10.1016/S0022-4405(01)00082-6.

Baker, L., & Scher, D. (2002). Beginning readers' motivation for reading in relation to parental beliefs and home reading experiences. *Reading Psychology*, 23(4), 239–269. https://doi.org/10.1080/713775283.

Baker, L., Scher, D., & Mackler, K. (1997). Home and family influences on motivations for reading. *Educational Psychologist*, 32(2), 69–82. https://doi.org/10.1207/s15326985ep3202_2.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B* (Methodological), 57(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bergman, P. (2019). How behavioral science can empower parents to improve children's educational outcomes. *Behavioral Science & Policy 5*(1), 52-67. https://doi:10.1353/bsp.2019.0004.

Bird, K. A., Castleman, B. L., Denning, J. T., Goodman, J., Lamberton, C., & Rosinger, K. O. (2019). Nudging at scale: Experimental evidence from FAFSA completion campaigns (No. w26158). National Bureau of Economic Research.

Cabell, S. Q., Zucker, T. A., DeCoster, J., Copp, S. B., & Landry, S. (2019). Impact of a parent text messaging program on pre-kindergarteners' literacy development. *AERA Open*, 5(1), 233285841983333. https://doi.org/10.1177/2332858419833339

Collins, L. M., Dziak, J. J., Kugler, K. C., & Trail, J. B. (2014). Factorial experiments: Efficient tools for evaluation of intervention components. *American Journal of Preventive Medicine*, 47(4), 498–504. https://doi.org/10.1016/j.amepre.2014.06.021

Cortes, K. E., Fricke, H., Loeb, S., Song, D. S., & York, B. N. (2019). Too little or too much? Actionable advice in an early-childhood text messaging experiment. *Education Finance and Policy*, 1–44. https://doi.org/10.1162/edfp_a_00304

Cortes, K., Fricke, H. D. U., Loeb, S., Song, D., & York, B. (2019). When behavioral barriers are too high or low – how timing matters for parenting interventions (No. w25964; p. w25964). National Bureau of Economic Research. https://doi.org/10.3386/w25964

DellaVigna, S., & Linos, E. (2020). RCTs to scale: Comprehensive evidence from two nudge units (No. w27594; p. w27594). National Bureau of Economic Research. https://doi.org/10.3386/w27594

Doss, C., Fahle, E. M., Loeb, S., & York, B. N. (2019). More than just a nudge: Supporting kindergarten parents with differentiated and personalized text messages. *Journal of Human Resources*, 54(3), 567–603. https://doi.org/10.3368/jhr.54.3.0317-8637R

Doss, Christopher, Hans Fricke, and Susanna Loeb. (2020). Math is for Girls: The Unequal Effects of Text Messaging to Help Parents Support Early Math Development. (EdWorkingPaper: 20-310). Annenberg Institute at Brown University. https://doi.org/10.26300/39zc-j672.

Hardy, M. (1993). Regression with dummy variables. SAGE Publications, Inc.

https://doi.org/10.4135/9781412985628

Hollands, F. M., Kieffer, M. J., Shand, R., Pan, Y., Cheng, H., & Levin, H. M. (2016). Cost-effectiveness analysis of early reading programs: A demonstration with recommendations for future research. Journal of Research on Educational Effectiveness, 9(1), 30–53. https://doi.org/10.1080/19345747.2015.1055639

Hurwitz, L. B., Lauricella, A. R., Hanson, A., Raden, A., & Wartella, E. (2015). Supporting Head Start parents: Impact of a text message intervention on parent–child activity engagement. Early Child Development and Care, 185(9), 1373–1389. https://doi.org/10.1080/03004430.2014.996217

Jacob, R. T., Armstrong, C., Bowden, A. B., & Pan, Y. (2016). Leveraging volunteers: An experimental evaluation of a tutoring program for struggling readers. *Journal of Research on Educational Effectiveness*, 9(S1), 67–92. https://doi.org/10.1080/19345747.2016.1138560

Kim, J. S., Asher, C. A., Burkhauser, M., Mesite, L., & Leyva, D. (2019). Using a sequential multiple assignment randomized trial (SMART) to develop an adaptive K–2 literacy intervention with personalized print texts and app-based digital activities. *AERA Open*. https://doi.org/10.1177/2332858419872701

Kim, J. S., Guryan, J., White, T. G., Quinn, D. M., Capotosto, L., & Kingston, H. C. (2016). Delayed effects of a low-cost and large-scale summer reading intervention on elementary school children's reading comprehension. *Journal of Research on Educational Effectiveness*, 9(sup1), 1–22. https://doi.org/10.1080/19345747.2016.1164780

Kraft, M. A., & Monti-Nussbaum, M. (2017). Can schools enable parents to prevent summer learning loss? A text-messaging field experiment to promote literacy skills. *The ANNALS*

of the American Academy of Political and Social Science, 674(1), 85–112. https://doi.org/10.1177/0002716217732009

Kugler, K. C., Trail, J. B., Dziak, J. J., & Collins, L. M. (2012). Effect coding versus dummy coding in analysis of data from factorial experiments. The Methodology Center, Pennsylvania State University. Retrieved from https://www.methodology.psu.edu/

Lynch, J., Anderson, J., Anderson, A., & Shapiro, J. (2006). Parents' beliefs about young children's literacy development and parents' literacy behaviors. *Reading Psychology*, 27(1), 1–20. https://doi.org/10.1080/02702710500468708

Marinak, B. A., Malloy, J. B., Gambrell, L. B., & Mazzoni, S. A. (2015). Me and My Reading Profile: A tool for assessing early reading motivation. *The Reading Teacher*, 69(1), 51–62. https://doi.org/10.1002/trtr.1362

Mayer, S. E., Kalil, A., Oreopoulos, P., & Gallegos, S. (2018). Using behavioral insights to increase parental engagement: The Parents and Children Together intervention. *Journal of Human Resources*, 0617-8835R. https://doi.org/10.3368/jhr.54.4.0617.8835R

Measure of Academic Progress. (2011). NWEA.

Muralidharan, K., Romero, M., & Wüthrich, K. (2019). Factorial Designs, Model Selection, and (Incorrect) Inference in Randomized Experiments (No. w26562). National Bureau of Economic Research. https://doi.org/10.3386/w26562.

Nickerson, D. W., & Rogers, T. (2010). Do you have a voting plan?: Implementation intentions, voter turnout, and organic plan making. *Psychological Science*, 21(2), 194–199. https://doi.org/10.1177/0956797609359326

North Carolina Department of Public Instruction. (2020). Beginning-of-Grade 3 Reading Test: North Carolina Test Specifications. https://www.dpi.nc.gov/media/10699/open.

Oettingen, G., & Reininger, K. M. (2016). The power of prospection: Mental contrasting and behavior change: Mental contrasting and behavior change. Social and Personality Psychology Compass, 10(11), 591–604. https://doi.org/10.1111/spc3.12271

Oreopoulos, P. (2020). Promises and Limitations of Nudging in Education. IZA Institute of Labor Economics: (Discussion Paper IZA DP No. 13718). https://ftp.iza.org/dp13718.pdf

Oreopoulos, P., Petronijevic, U., Logel, C., & Beattie, G. (2020). Improving non-academic student outcomes using online and text-message coaching. *Journal of Economic Behavior & Organization*, 171, 342–360. https://doi.org/10.1016/j.jebo.2020.01.009

Robinson, C. D., Lee, M. G., Dearing, E., & Rogers, T. (2018). Reducing student absenteeism in the early grades by targeting parental beliefs. *American Educational Research Journal*, 55(6), 1163–1192. https://doi.org/10.3102/0002831218772274

Rogers, T., & Feller, A. (2018). Reducing student absences at scale by targeting parents' misbeliefs. *Nature Human Behaviour*, 2(5), 335–342. https://doi.org/10.1038/s41562-018-0328-1

Rogers, T., Duncan, T., Wolford, T., Ternovski, J., Subramanyam, S., & Reitano, A. (2017). A randomized experiment using absenteeism information to "nudge" attendance REL 2017-252. *Regional Educational Laboratory Mid-Atlantic*. http://ies.ed.gov/ncee/edlabs

Rogers, T., Milkman, K. L., John, L. K., & Norton, M. I. (2015). Beyond good intentions: Prompting people to make plans improves follow-through on important tasks. *Behavioral Science & Policy*, 1(2), 33–41. https://doi.org/10.1353/bsp.2015.0011

Schanzenbach, D. Whitmore. (2006). What have researchers learned from project STAR? Brookings Papers on Education Policy, 2006(1), 205–228.

https://doi.org/10.1353/pep.2007.0007

Smythe-Leistico, K., & Page, L. C. (2018). Connect-text: Leveraging text-message communication to mitigate chronic absenteeism and improve parental engagement in the earliest years of schooling. *Journal of Education for Students Placed at Risk (JESPAR)*, 23(1–2), 139–152. https://doi.org/10.1080/10824669.2018.1434658

Somers, M., Collins, L., Maier, M. (2014). *Review of Experimental Designs for Evaluating Component Effects in Social Interventions*. Produced by American Institutes for Research for *Head Start Professional Development: Developing Evidence for Best Practices in Coaching*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation. Retrieved from https://www.mdrc.org.

Sonnenschein, S., Baker, L., Serpell, R., Scher, D., Fernandez-Fein, S., & Munsterman, K. A. (1996). Strands of emergent literacy and their antecedents in the home: Urban preschoolers early literacy development (No. 48; p. 49). National Reading Research Center.

Sonnenschein, S., Baker, L., Serpell, R., Scher, D., Truitt, V. G., & Munsterman, K. (1997). Parental beliefs about ways to help children learn to read: The impact of an entertainment or a skills perspective. *Early Child Development and Care*, 127(1), 111–118. https://doi.org/10.1080/0300443971270109

Sonnenschein, S., Baker, L., Serpell, R., & Schmidt, D. (2000). Reading is a source of entertainment: The importance of the home perspective for children's literacy development. In *Play and literacy in early childhood: Research from multiple perspectives* (pp. 107–124). Lawrence Erlbaum Associates Publishers.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

York, B. N., Loeb, S., & Doss, C. (2018). One step at a time: The effects of an early literacy text messaging program for parents of preschoolers. *Journal of Human Resources*, 0517-8756R. https://doi.org/10.3368/jhr.54.3.0517-8756R

Figure 1

Theory of Change for Repeated Family Messaging Interventions

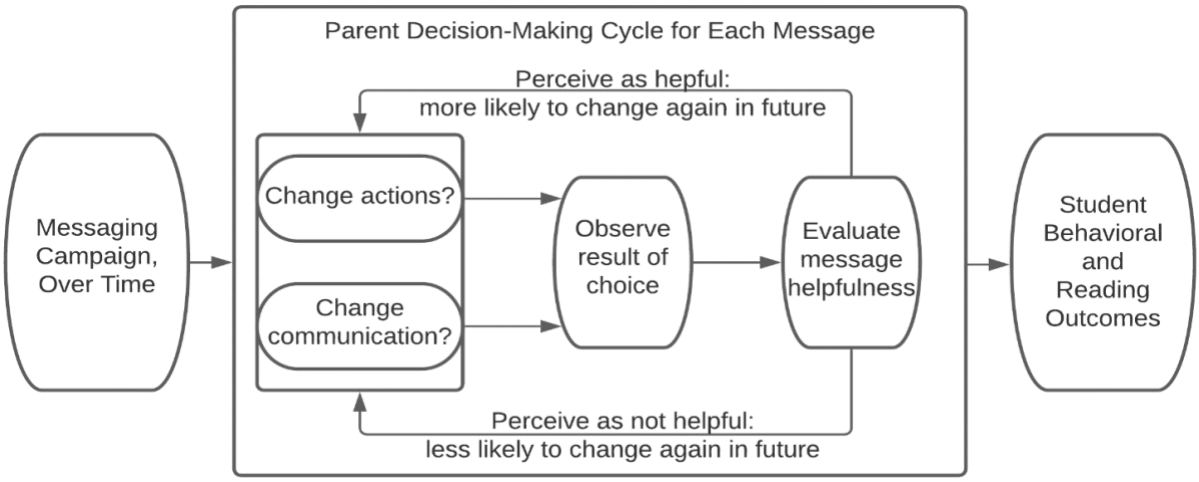


Table 1

Baseline characteristics and balance checks

	Full Sample	Personalization vs. Not Difference	Entertainment vs. Instrumental Difference	Combination vs. Instrumental Difference	Goal Setting vs. Not Difference
White (%)	0.184	0.004	-0.000	-0.008	-0.001
	(0.388)	(0.009)	(0.012)	(0.010)	(0.009)
Black (%)	0.386	0.008	-0.023	-0.011	0.008
	(0.487)	(0.012)	(0.018)	(0.017)	(0.012)
Hispanic (%)	0.317	-0.008	0.022	0.001	-0.004
	(0.465)	(0.011)	(0.017)	(0.017)	(0.013)
Asian (%)	0.079	-0.001	0.008	0.023*	0.003
	(0.270)	(0.008)	(0.012)	(0.011)	(0.008)
Limted English Proficiency	0.226	-0.016	0.014	-0.000	0.013
	(0.418)	(0.010)	(0.012)	(0.015)	(0.012)
Low SES	0.405	0.017+	0.004	-0.004	-0.004
	(0.491)	(0.010)	(0.012)	(0.012)	(0.009)
Med SES	0.385	-0.000	0.002	0.011	0.004
	(0.487)	(0.010)	(0.015)	(0.014)	(0.011)
High SES	0.205	-0.018*	-0.009	-0.010+	-0.001
	(0.404)	(0.009)	(0.009)	(0.006)	(0.010)
Standardized Spring MAP ELA RIT	-0.000	0.040+	0.001	-0.014	-0.022
	(0.988)	(0.024)	(0.035)	(0.035)	(0.025)
Standardized Spring MAP Math RIT	-0.000	0.052+	0.001	0.005	-0.042
	(0.988)	(0.027)	(0.033)	(0.038)	(0.028)
N	5175	4678	3119	3128	4678

Source: district administrative records

Notes: Each row represents a separate regression. Point estimates reflect the condition-reference group differences derived from a model that includes indicators for the randomization block. Robust standard errors clustered at the school-grade level (in parentheses). P-values reflect inference prior to corrections for multiple hypothesis testing.
+p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table 2

Differential Effects of Text Messaging Components on Parent Outcomes

	Observed reading behaviors in app			Self-reported behaviors & beliefs	
	Ever logged in (%)	Minutes on app	Books completed	Frequency of reading activities	Found texts helpful (%)
Panel A - Main Effects					
Personalization vs. Not	0.028*	1.626**	0.681**	0.046	0.023
	(0.012)	(0.560)	(0.254)	(0.144)	(0.163)
Entertainment vs. Instrumental	0.004	-0.386	0.033	-0.273+	-0.090
	(0.020)	(1.016)	(0.434)	(0.160)	(0.272)
Combination vs. Instrumental	-0.020	-0.179	-0.345	0.249	0.360
	(0.020)	(0.891)	(0.397)	(0.191)	(0.243)
Goals vs. Not	-0.013	-0.245	-0.187	0.131	0.318+
	(0.013)	(0.675)	(0.287)	(0.156)	(0.169)
Panel B - Two-Way Interaction Effects					
Personalization x Entertainment	-0.013	-0.282	-0.149	0.191	0.211
	(0.019)	(1.010)	(0.435)	(0.150)	(0.234)
Personalization x Combination	0.006	-0.116	-0.187	-0.255	-0.097
	(0.016)	(0.948)	(0.391)	(0.183)	(0.225)
Personalization x Goals	0.014	0.365	0.138	-0.044	-0.243
	(0.012)	(0.587)	(0.234)	(0.128)	(0.176)
Entertainment x Goals	-0.021	-0.947	-0.464	-0.089	0.058
	(0.017)	(0.855)	(0.337)	(0.250)	(0.178)
Combination x Goals	0.041*	1.858*	0.686+	0.129	0.280
	(0.020)	(0.945)	(0.395)	(0.176)	(0.196)
N	4678	4678	4678	319	266

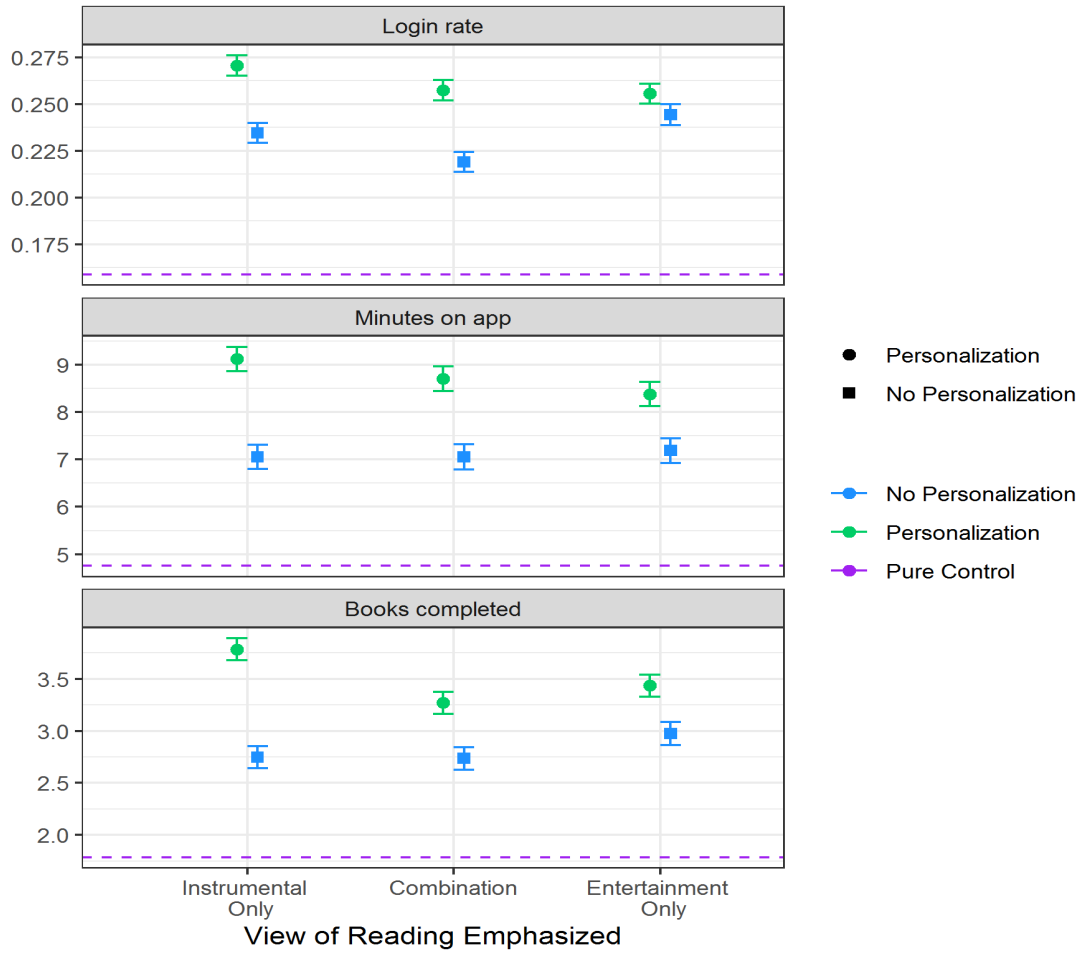
Source: app-use records, survey of parent subsample, district administrative records

Notes: Point estimates derived from effect-coded regressions that include all treatment factors and their interactions, as well as the following covariates: gender, race/ethnicity, participation in gifted program, participation in Special Education, English learner status, neighborhood SES, language of text messages, baseline reading and math scores, and indicators of the randomization block. Robust standard errors clustered at the school-grade level (in parentheses). P-values reflect inference prior to corrections for multiple hypothesis testing.

+p<0.10, *p<0.05, **p<0.01, ***p<0.001

Figure 2

Predicted Reading Behaviors, by Personalization and View-of-Reading Conditions



Source: app use data, district administrative records

Notes: Predicted values and confidence intervals for the six groups are calculated from model parameters and aggregated by the personalization and view-of-reading conditions, collapsing across the view-of-reading condition. The pure control value is observed directly from the data.

Table 3

Differential Effects of Text Messaging Components on Student Test Scores

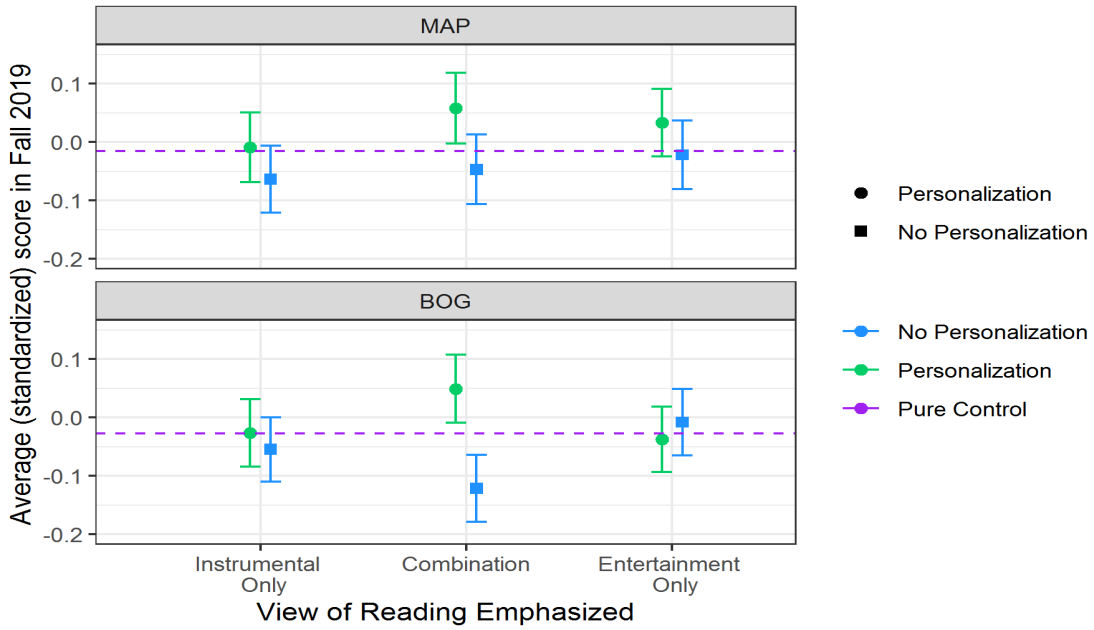
	MAP	Beginning of Grade
Panel A - Main Effects		
Personalization vs. Not	0.034*	0.019
	(0.016)	(0.021)
Entertainment vs. Instrumental	0.028	0.021
	(0.025)	(0.033)
Combination vs. Instrumental	0.026	-0.008
	(0.022)	(0.032)
Goals vs. Not	0.007	0.007
	(0.016)	(0.027)
Panel B - Two-Way Interaction Effects		
Personalization x Entertainment	-0.011	-0.078*
	(0.028)	(0.037)
Personalization x Combination	0.025	0.106***
	(0.027)	(0.032)
Personalization x Goals	-0.005	-0.010
	(0.016)	(0.020)
Entertainment x Goals	-0.009	0.016
	(0.024)	(0.040)
Combination x Goals	-0.010	-0.029
	(0.021)	(0.042)
N	3961	2039

Source: district administrative records

Notes: Point estimates derived from effect-coded regressions that include all treatment factors and their interactions, as well as the following covariates: gender, race/ethnicity, participation in gifted program, participation in Special Education, English learner status, neighborhood SES, language of text messages, baseline reading and math scores, and indicators of the randomization block. Robust standard errors clustered at the school-grade level (in parentheses). P-values reflect inference prior to corrections for multiple hypothesis testing.
+p<0.10, *p<0.05, **p<0.01, ***p<0.001

Figure 3

Predicted Test Scores, by Personalization and View-of-Reading Conditions



Source: district administrative records

Notes: Predicted values and confidence intervals for the six groups are calculated from model parameters and aggregated by the personalization and view-of-reading conditions, collapsing across the view-of-reading condition. The pure control value is observed directly from the data.

Table 4

Differential Effects of Text Messaging Components on Alternate Student Outcomes

	Student Self-Reports		Qualitative Experience Among App Users	
	Proportion of books read (%)	Reading motivation (MMRP)	Enjoyed activities	Felt like a good reader
Panel A - Main Effects				
Personalization vs. Not	0.025*	0.001	0.023	-0.003
	(0.013)	(0.034)	(0.060)	(0.068)
Entertainment vs. Instrumental	-0.019	0.005	-0.026	-0.125
	(0.016)	(0.057)	(0.097)	(0.098)
Combination vs. Instrumental	0.018	-0.078+	0.079	0.115
	(0.017)	(0.046)	(0.087)	(0.082)
Goals vs. Not	0.002	0.005	-0.050	-0.031
	(0.010)	(0.029)	(0.061)	(0.068)
Panel B - Two-Way Interaction Effects				
Personalization x Entertainment	-0.001	0.039	0.114	0.191*
	(0.013)	(0.047)	(0.120)	(0.097)
Personalization x Combination	-0.003	0.002	-0.050	-0.154
	(0.017)	(0.055)	(0.115)	(0.105)
Personalization x Goals	-0.011	0.046+	0.058	0.052
	(0.009)	(0.024)	(0.066)	(0.067)
Entertainment x Goals	-0.031+	0.102*	0.074	0.054
	(0.016)	(0.047)	(0.085)	(0.095)
Combination x Goals	0.010	-0.125*	-0.134	-0.073
	(0.017)	(0.050)	(0.098)	(0.101)
N	3472	3490	995	995

Source: student survey, district administrative records, app-use records

Notes: Point estimates derived from effect-coded regressions that include all treatment factors and their interactions, as well as the following covariates: gender, race/ethnicity, participation in gifted program, participation in Special Education, English learner status, neighborhood SES, language of text messages, baseline reading and math scores, and indicators of the randomization block. Robust standard errors clustered at the family level (in parentheses). P-values reflect inference prior to corrections for multiple hypothesis testing.

+p<0.10, *p<0.05, **p<0.01, ***p<0.001

Figure S.1

Consort Diagram

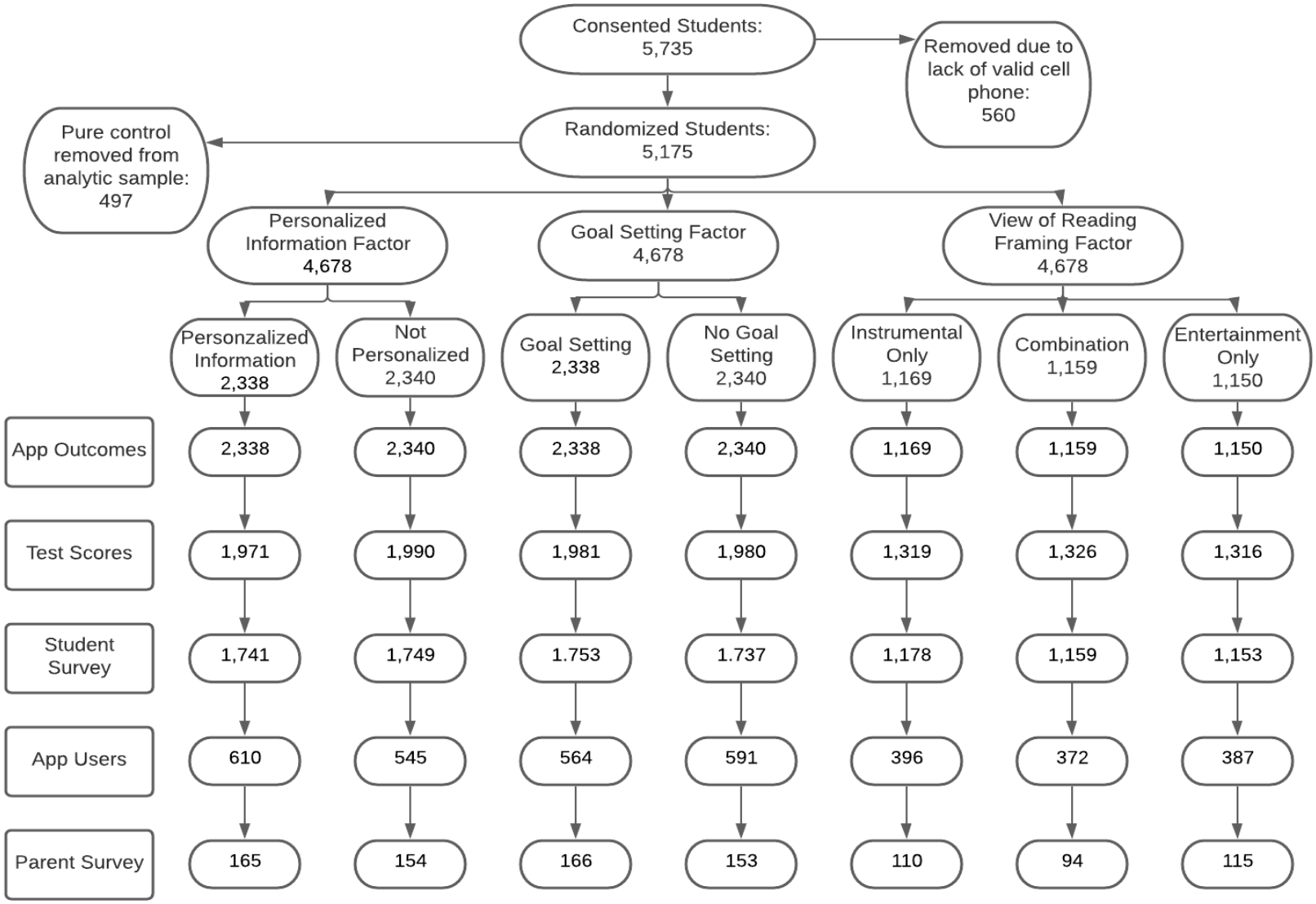


Table S.1

Text Message Topics, Organized by Theme

Summer Reading Resources	**Reminders to Engage in Reading Activities**	**Monitoring progress**
· App feature: availability of books · App feature: personalized activities/goal reminder · App feature: "catching words" · Resource: chromebooks available at library · App feature: new content · App feature: 2 sets of activities	· Tip: talking about books · App is great · Tip: planning time to read · Social pressure: number of users · Tip: talk about favorite book	· Kickoff & goal setting · Monitoring: check-in on goals/progress · Monitoring: check-in on goals/progress · Closeout & survey preview

Table S.2

Comparing Message Variations Across Conditions for Two Example Messages

	Generic	**Personalized information**
Message 1		
Instrumental view	The MORE@Home app contains personalized activities for each book that will help [StudentFirstName] **develop reading skills**.	The MORE@Home app contains personalized activities for each book. [StudentFirstName] can **practice different reading skills** for both **[Book1Title] and [Book2Title]**.
Entertainment view	The MORE@Home app contains personalized activities for each book. We think [StudentFirstName] **will have fun doing them all!**	The MORE@Home app contains personalized activities for each book. [StudentFirstName] can **have fun exploring them** for both **[Book1Title] and [Book2Title]**.
Message 2		
No goals	We are already 6 weeks into summer vacation! Hopefully you and [StudentFirstName] are making progress on your summer reading list!	We are 6 weeks into summer vacation and [StudentFirstName] **has used [NumBooksAccessed] books on the MORE@Home app**. Keep reading to get through all of them!
Goals	We're already 6 weeks into summer vacation! At this point, you should be about **1/2 of the way to your summer reading goal**.	We're already 6 weeks into summer vacation! At this point, you should be about 1/2 of your way to your **summer reading goal of [BookGoal] books**.

Table S.3

Treatment Conditions in Factorial Design with Effect Coding

Group	Received Text Messages	Value	Goals Condition	Value	Personalization Condition	Value	Reading Value Condition	Value (Entertain)	Value (Both)
1	No	0		.		.		.	.
2	Yes	1	No Goals	-1	No Personalization	-1	Instrumental Only	-1	-1
3	Yes	1	No Goals	-1	No Personalization	-1	Entertainment Only	1	0
4	Yes	1	No Goals	-1	No Personalization	-1	Both	0	1
5	Yes	1	No Goals	-1	Personalization	1	Instrumental Only	-1	-1
6	Yes	1	No Goals	-1	Personalization	1	Entertainment Only	1	0
7	No	1	No Goals	-1	Personalization	1	Both	0	1
8	Yes	1	Goals	1	No Personalization	-1	Instrumental Only	-1	-1
9	Yes	1	Goals	1	No Personalization	-1	Entertainment Only	1	0
10	Yes	1	Goals	1	No Personalization	-1	Both	0	1
11	Yes	1	Goals	1	Personalization	1	Instrumental Only	-1	-1
12	Yes	1	Goals	1	Personalization	1	Entertainment Only	1	0
13	Yes	1	Goals	1	Personalization	1	Both	0	1

Table S.4

Tests of differential attrition across data sources and experimental factors

	Personalization vs. Not			Entertainment vs. Instrumental			Combination vs. Instrumental			Goal Setting vs. Not		
	Component Retention	Reference Group Retention	Diff. (s.e.)	Component Retention	Reference Group Retention	Diff. (s.e.)	Component Retention	Reference Group Retention	Diff. (s.e.)	Component Retention	Reference Group Retention	Diff. (s.e.)
MAP Assessment	0.85	0.84	0.01	0.85	0.84	0.01	0.85	0.84	0.01	0.85	0.85	0.00
			(0.01)			(0.01)			(0.01)			(0.01)
Student Survey	0.73	0.73	0.00	0.73	0.74	0.00	0.72	0.74	-0.01	0.73	0.73	0.01
			(0.01)			(0.02)			(0.02)			(0.01)
Parent Survey	0.07	0.07	0.00	0.07	0.07	0.00	0.06	0.07	-0.01	0.07	0.06	0.01
			(0.01)			(0.01)			(0.01)			(0.01)
App Users	0.26	0.23	0.03 **	0.25	0.25	0.00	0.24	0.25	-0.01	0.24	0.25	-0.01
			(0.01)			(0.02)			(0.02)			(0.01)
Baseline N	2340	2336		1548	1570		1558	1570		2339	2337	