Economics
Working Papers
2023-01

Supporting Language Development through a Texting Program: Initial Results from Denmark

Susanna Loeb, Michala Iben Riis-Vestergaard and Marianne Simonsen

AARHUS BSS
DEPARTMENT OF ECONOMICS AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

Supporting Language Development through a Texting Program: Initial Results from Denmark

Susanna Loeb
Center for Education Policy Analysis
Stanford

Michala Iben Riis-Vestergaard
MIT

Marianne Simonsen
Department of Economics and Business Economics
Aarhus University

January 2023

This paper presents the results from a large-scale family-level randomized controlled trial evaluation of a texting program, TipsbyText, which had as its primary goal to enhance child language development through supporting parents and caregivers in creating playful, language stimulating activities. The program was delivered to parents of children aged 3-6 in Danish preschools and the evaluation combines the data from the randomized controlled trial with population-level register-based data. We measure outcomes immediately after program delivery and at one-year follow-up. While the program was generally well liked by parents and delivered as intended with relatively low dropout, TipsByText did not affect children's language development, neither for the full sample, nor for any pre-defined subgroups. We discuss possible reasons for the lack of positive effects.

Keywords: Text messages; randomized trial; language development; children.

JEL codes: I2

Acknowledgements: Parts of this manuscript were included in Michala Riis-Vestergaard's PhD dissertation: "TipsByText: A Parent-Oriented Language Intervention for Preschool Children" presented to the faculty of Princeton University in candidacy for the degree of Doctor of Philosophy in 2021 (Riis-Vestergaard, M., 2021). We are grateful to Johanne Louise Elholm Bergmann who led the implementation of TipsByText. We also appreciate outstanding research assistance from Signe Haahr Petersen and Mikkel Stahlschmidt, and insightful comments from Maria Knoth Humlum, Rasmus Landersø, and Zazie Tolmer. The project has been registered with the Social Science Registry: https://www.socialscienceregistry.org/trials/6393, RCT ID: AEARCTR-0006393. The usual disclaimer applies.

1. Introduction

A rich literature points to the importance of early investments and circumstances for children's long-run academic and behavioral outcomes; and many studies point to the family and interactions within the family as crucial drivers for the development of children's skills (e.g., Carneiro and Heckman, 2003; Almond, Currie and Duque, 2018). Yet, despite the promise of parent-oriented interventions for improving early life skills, there is limited causal evidence on the effectiveness of scalable parent-oriented interventions. This paper exploits a randomized controlled trial to investigate the effects of a texting program, TipsByText, which focused on child language development and had as its primary goals to support parents and caregivers in creating playful, language stimulating activities.

In practice, the texting program was delivered to parents of children aged 3-6 in Danish preschools and the evaluation combines the data from the randomized controlled trial with population-level register-based data. To reach as many parents as possible, implemented the evaluation with an opt-out design. This meant that all parents of children enrolled in the participating preschools received information about the experiment. Eligible children whose parents did not opt-out were subsequently randomized to either receive texts or to the control group. A first key result is that the opt-out design did support broad outreach: in total, 3,611 out of 3,804 children (95%) were included in the randomization. We find, however, that though the program was delivered as intended with relatively low dropout, TipsByText did not affect children's language development as measured by their language assessment scores, neither shortly after program delivery ended, nor in the one-year follow-up. Our main estimates are close to zero and precisely estimated. We do not detect meaningful, statistically significant effects on any pre-specified subgroups.

There are essentially two places where the theory of change can break down: either the intervention failed to affect parents' behavior, or the parental behavior change did not, in turn, improve the children's language skills. A consecutive qualitative evaluation of TipsByText found that the program was generally well liked and used by parents (Tolmer, 2021), indicating that the first step of the theory of change is likely to hold. Thus, a possible explanation for the lack of results is a disconnect between the tasks that TipsByText encouraged and the literacy needs of the population of children in this study.

Our paper contributes to a very small literature that causally explores the link between texting programs targeted at parents and child literacy. Our work relates most closely to that of York, Loeb and Doss (2019) who use a randomized experiment to evaluate the effects of Ready4K!, a parent

oriented text message based program lasting eight months. The overarching goal of their program was to support the literacy, math, and socioemotional development of the participating children. During the first year of the intervention (the 2013-2014 school year), treated families received text messages that were solely concerned with children's literacy skills. During the second year of the intervention (the 2015-2016 school year), in contrast, another set of families received text messages that addressed literacy skills but also mathematics and socioemotional skills. Parents in their treatment group received three texts messages per week during the school year, with each week addressing a particular set of skills. Importantly, the literacy component of Ready4K! served as a prototype for our program, though suitably adapted to the Danish context. Ready4K! was delivered to randomly selected parents of four-year-old preschoolers in the lowest third of the income distribution among San Francisco's preschool market. Parents were recruited through the district's existing enrollment practices; enrollment clerks were offered ten dollars for each family they enrolled, and the first cohort of parents were offered monetary incentives to enroll and to stay enrolled. The scale of the experiment was smaller than ours with 1,031 of 1,761 eligible families (58%) enrolled in the study. They find strong evidence that parents in the treatment group used the content of the text messages and found the program to be helpful. Moreover, they find that the first year of the experiment generated small but insignificant positive estimates (0.06 SD) in an early literacy assessment. In the second year, in contrast, there is a robust 0.146 SD ($p < 0.05$) increase in the average of all questions. Thus, an outstanding question is still whether the literacy portion of the program tested in the first year can significantly improve early language skills. We evaluate a version of this program with a larger sample.

Apart from the slightly different content of the treatment and the scale of the study, a key difference between our design and theirs lies in the recruitment strategy. While parents in York et al. (2019) opted into the program, parents in our set-up had to opt out actively, which involved overcoming a small barrier (making a phone call or writing an email). One benefit of an opt-in design is that all participants have revealed an interest in participating. However, a weakness is that it might not reach everyone who is interested in participating, e.g., because of lack of information or lack of bandwidth to sign up. Thus, a program that requires participants to opt in might not reach some of the families who would benefit most from the program; the benefit of an opt-out design is to make sure that everyone who could be interested in participating are reached. It is also possible that the population under study in York et al. (2019) is considerably worse off in terms of their socio-economic

background – and therefore likely in terms of parental investment behaviors and children's literacy needs too – when compared to our Danish equivalent.

Other related work is Mayer et al. (2019). Their intervention, The Parents and Children Together, provided tablets to parents of children in Head Start and set goals for children's reading at home. These goals were aided by text messages. Their treatment included three behavioral tools (a commitment device, reminders via text messages, and a social incentive) plus information on the importance of parents reading to their children. This study found a substantial increase in parent-reported home-based reading but did not study children's academic outcomes; our study includes detailed language development measures on all children both immediately after the program delivery and at one-year follow-up.

We also speak to the literature that is concerned with family-based home literacy interventions more broadly; see e.g., Reese, Sparks and Leyva (2010) for a review. Here, there are in general two categories of parent-oriented interventions that have been studied rigorously and have been found to have consistent positive effects on children's early language skills: Shared-reading interventions and parent-training interventions. However, both types of interventions come with their own set of strengths and limitations. A strength of the shared reading interventions is that they are typically rather cheap and easy to scale (e.g., sending books to families), however, a limitation is that they focus only on one aspect of parent-child verbal interactions, shared reading, which must be assumed to constitute only a fraction of overall parent-child verbal interactions. Parent education programs have the strength that they typically cover many aspects of children's early development, but its weakness is that they are often costly to implement and difficult to scale. Thus, there is often a trade-off between scope and scalability in parent-oriented interventions. Furthermore, many studies of parent-oriented interventions are based on very small sample sizes. E.g., in the 11 studies included in the review of Reese, Sparks and Leyva (2010), the largest study is based on 248 families. Our project contributes to this literature by addressing the problems of (1) a trade-off between scope and scalability in parent-oriented interventions, and (2) small sample sizes in related studies. We do this by testing the impact of a parent-oriented intervention which (1) is cheap to implement, (2) is easy to scale, (3) has a broad scope aimed at impacting a wide variety of parent-child verbal interactions, and (4) is evaluated in a naturalistic setting with a large and diverse sample.

Andersen and Nielsen (2016), for example, set up a large-scale randomized field trial where parents receiving a reading intervention were told about the malleability of their child's reading abilities and

how to support their child by praising his/her effort rather than his/her performance. This intervention increased the reading and writing achievements of all participating children, including immigrant children with non-Western backgrounds and children with low-educated mothers.

We structure the remainder of the paper as follows: Section 2 provides details about the intervention and the adaptation of the Ready4K! to the Danish context. Section 3 presents our recruitment strategy and randomization details, Section 4 presents our data, and Section 5 presents the results of the evaluation. Finally, Section 6 discusses possible explanations behind our results and Section 7 concludes.

2. The TipsByText program

TipsByText is a text-messaging program that aims to make it easier for parents to support their children's language development in the preschool years (ROCKWOOL Foundation, 2019). For eight months, parents receive three weekly (Monday, Wednesday, and Friday) text messages with tips for fun and simple language-stimulating games and activities they can do with their children. The text messages are formulated in non-complex language and the tone in the texts is short, direct, and with a call for immediate action in daily routines, e.g., bath time, mealtime or when picking the child up from preschool. The curriculum is generally aimed at parents of preschool-aged children, ages 3-5, and the baseline language is Danish although parents have the option to switch to several other languages identified as the most commonly spoken languages in our collaborating municipalities.[1]

TipsByText was adapted to the Danish context from the American parent education program, Ready4K, which was developed in 2013 by Susanna Loeb, Professor of Education and International and Public Affairs at Brown University, and Benjamin York, Founder and CEO of ParentPowered.[2] TipsByText, as well as its American counterpart, READY4K!, is designed around three core principles: First, the activities should be fun and easy. Second, lecturing is to be avoided; it is at the parents' discretion how they use the text messages, and no one monitors how and how much they do so. Finally, practice makes perfect: the text messages should be based on and promote a growth mindset (Dweck, 2007).

Furthermore, TipsByText is designed to address behavioral barriers to involved parenting. Some parents may have insufficient information about the impact their child-directed speech has on their

[1] The default language of the text messages was Danish, but the parents could switch language to English, Arabic, Turkish, Ukrainian, Polish, Romanian, or Urdu. Only 1.1% of parents switched language.
[2] See https://ready4k.parentpowered.com for more information

child's language development. To overcome this barrier, parents receive a "FACT" text each Monday designed to generate knowledge by highlighting the importance of particular skills and the parent's role in developing that skill. In addition, parenting is a complex task, and for some parents the cognitive load of parenting can result in a low prioritization of active language stimulation. To help alleviate this barrier, the parents receive a "TIP" text each Wednesday, which contains a specific and creative exercise in how to incorporate a language-stimulating element into a daily routine that is already taking place. Third, to motivate parents and keep them engaged in the language support of their children, the parents receive a "GROWTH" text each Friday, which provides encouragement and a deeper insight into the benefits of the exercise given on Wednesday. An example of a week of text messages is as follows:

- FACT (Monday): Children need to know that letters make up words. Research shows that kids with good letter knowledge become good readers.
- TIP (Wednesday): Point out the letters in your child's name in magazines, on signs & at the store. Have your child try. Who can find the most?
- GROWTH (Friday): Keep pointing out letters. You're making a big effort to support your child's learning. Now, when you point out a letter ask: what sound does it make?

2.1 Adapting the program to the Danish context

TipsByText was adapted to the Danish context by The ROCKWOOL Foundation in collaboration with Susanna Loeb (one of the original designers of the program), Danish language researcher and consultant Pia Thomsen, and language consultants, parents, and preschool teachers. Though one goal of the adaptation process was to change the curriculum and the wording of the text messages as little as possible, the adaptation involved two general changes (in addition to translating the texts into Danish): an adaptation to the Danish learning culture, and an expansion of the curriculum to focus on more domains of early language development than early literacy. The core design principles were preserved.

The US-based intervention, READY4K!, is based on the American learning culture, which emphasizes formal learning and a "readiness for school" approach from a young age (OECD, 2006). In this view, the primary purpose of preschool is to prepare children for primary school. Thus, cognitive development and the acquisition of knowledge and skills is at the center of the preschool curricula. One way the "readiness for school" approach to early learning is visible in the READY4K!

curriculum is that all GROWTH texts in the US curriculum are structured as: "By [taking up the activity of the week], you're preparing your child 4K!" (York et al., 2019, online appendix). The Danish view of learning, in contrast, comes from the social pedagogy tradition, which emphasizes broad developmental goals, such as socio-emotional development, personal and social skills, artistic and cultural development and learning through lived situations (OECD, 2006). In this view, the role of childcare of 3–5-year-olds is not to prepare the children for primary school but to provide the children with care, socialization, and a broad preparation for life. In Danish preschools, as well as in children's homes, learning happens primarily through play, interaction, activity, and personal investigation. As research conducted in Danish preschools has found that both parents and pedagogues dislike working with academic activities and prefer activities that focus on interrelations and interaction, e.g., play-based activities (Nielsen et al., 2017), the content of the text messaged was adapted to match the Danish learning culture. The adapted curriculum still focuses on promoting language development and early literacy but where relevant the text messages encourage parents to promote this development through playful activities instead of academic and cognitive activities. This adaption is supported by research showing that interaction, play and dialogue can have a positive effect on children's language development (e.g., Heidlage et al., 2018).

Whereas the US curriculum focuses primarily on supporting early literacy (letters, reading direction, etc.), the Danish curriculum has been expanded to focus on two additional components of early language development: vocabulary (formal language; understanding and producing words), and language use/communication (conversation, dialogue, narrativity, taking turns, and initiatives). These dimensions were included to match the focus of the language assessment tool, Sprogvurdering 3-6, which was developed based on rigorous research on early indicators of language development (Ministry of Children and Education, 2019).

3. Experimental design

The experiment is designed as a family-level randomized controlled trial with an opt-out design. The RCT is funded by The ROCKWOOL Foundation and carried out in collaboration with five partnering municipalities in Denmark: Hoeje-Taastrup, Ikast-Brande, Langeland, Lejre, and Middelfart.

3.1 Recruitment and sample selection

Municipalities were recruited by broad outreach. Eight municipalities agreed to participate in the RCT, but three municipalities opted out before the start of the RCT due to the increased workload associated with participation (collecting data on parent's phone numbers, conducting extra language assessments, and distributing information material to parents). The five collaborating municipalities vary in size – from just above 12,000 inhabitants in Langeland municipality in 2020 to more than 50,000 inhabitants in Hoeje-Taastrup municipality – and the municipalities are distributed geographically across the country.

The collaborating municipalities decided which public preschools would participate in the RCT; in four out of five collaborating municipalities, all public preschools participated, and in the largest municipality, Hoeje-Taastrup, 2/3 of public preschools participated. The participating preschools in Hoeje-Taastrup were typically selected from lower socioeconomic areas, and families from lower socioeconomic strata are thus slightly overrepresented in this municipality. 3,804 children ("the preschool sample") born between 2013 and 2016 were enrolled in the participating preschools on October 1, 2019.

3.2 Information and timeline

Figure 2 presents the timeline of the study. Before the randomization was conducted, all parents to children in the participating preschools were informed about the RCT. Information was provided via a pamphlet that was delivered in printed form to the parents of all children in each preschool by the preschool staff.[3] The pamphlet describes the intervention, the evaluation (including the randomization), the data collection, the research team, and the opt-out design of the intervention. It also included a phone number and email the parents could contact with questions or if they wanted to opt out of the evaluation.

The preschool teachers were informed about the program from their supervisors, and each preschool received an extended version of the information pamphlet. The teachers were responsible for entering parental phone numbers into a database and for carrying out language assessments of the children before the start of the program (baseline), immediately after the end of the program (endline 1), and one year after the end of the program (endline 2). The language assessment is explained in detail in section 4. Except for their role in data collection, the teachers did not play any role in the program

[3] To increase the likelihood that the information reached all parents, the pamphlet was also uploaded to most preschools' intranet. It was translated into eight languages; all parents received the printed version in Danish, and for each child, the personnel in the preschool decided whether to attach a translated version.

delivery or evaluation, and they were explicitly instructed to not start conversations with the parents about the program so as to keep the program separate from the preschool.

Children were included in the randomization if three conditions were met on January 1, 2020: (1) the parents had not opted out of the trial, (2) the municipality had provided at least one parental phone number, and (3) the preschool teachers deemed the child fit to participate in the language assessment. In total, 3,611 children were included in the randomization and, importantly, only 52 parents actively opted out before the start of the program. The age distribution of the children included in the trial was as follows: 1,028 children born in 2016 (age 3 at the start of the intervention), 1,142 children born in 2015 (age 4), 1,242 children born in 2014 (age 5), and 92 children born in 2013 (age 6 with delayed school start).[4]

3.3 Randomization

We randomize treatment status at the family level to ensure that everyone within a family received the same treatment.[5] We conducted the randomization using the randtreat command in Stata 15.1. We subdivide families into three different groups: an *interview group* that were to participate in a concurrent qualitative evaluation, a *treatment group*, and a *control group*. The former group consisted of 120 families (with 130 children) who were drawn at random (stratified by municipality and whether the youngest child's baseline language assessment score was above or below the median).[6] We subsequently removed the *interview group* from the main RCT sample because we were worried that being interviewed at depth about their use of the intervention could affect their engagement with the intervention and thus its effects. We discuss our findings in light of the qualitative evaluation in Section 6 below. We randomized the remaining families 50/50 into the treatment and control groups stratified by three variables:[7]

a) The youngest child's preschool (70 strata)

[4] Age is missing for seven children. The 6-year-old children with delayed school start were included in the RCT for implementation purposes to ensure consistency in implementation across the preschools.
[5] At the time of randomization, we did not have data on which children were siblings. To minimize spillovers, we therefore categorized all children who shared at least one parental phone number as siblings. Thus, we might have categorized some stepchildren who shared a parent but did not live together as belonging to the same family. In practice, to ensure that we assigned all children within a family the same treatment status, randomization was performed for the younger siblings in the sample. We subsequently assigned the oldest siblings the same treatment status.
[6] The qualitative evaluation was to be based on 40 interviews, and the number 120 was determined with an assumption that 1/3 of families would accept the invitation to be interviewed. In the end, the randomization process assigned only 119 families to this group.
[7] In total, there were 840 strata. We allocated misfits using the global method in the randtreat command, meaning that misfits were assigned in equal proportions to the treatment and control group.

b) Whether the youngest child scored above/below the median within their age group in the baseline language assessment, or whether their baseline language assessment was missing (3 strata)
c) The youngest child's age on November 1, 2020 (4 strata)

In total, 1,743 children were randomized into the treatment group and 1,734 children were randomized into the control group. Figure 1 presents a flow chart of the experimental design.

To minimize bias, we aimed to minimize the probability that the preschool staff, who conducted the language assessments, knew the allocation of children into the treatment or control group. We used three strategies to achieve this goal. First, the preschool staff was given no information about the allocation. Second, the preschool staff was explicitly instructed not to initiate conversations with the parents about the intervention. Third, to minimize the risk of parents engaging the preschool staff in conversation about the intervention, the informational pamphlet did not mention the preschool. Instead, the sender of the pamphlet was the municipality and the ROCKWOOL Foundation Interventions Unit.

4. Data

Our main analyses make use of three key data sources with individual level information about children and their families. Crucial for our study is information from the Danish language assessment tool, "Sprogvurdering 3-6." The tool is designed to measure the language skills of children in preschool (age 3 to 6) and first year of primary school (ages 5 or 6). The language assessment is widely used in preschools and primary schools as a screening device to identify which children need to be referred to a language specialist or to be monitored for language delays (Bleses et al., 2010). We measure language development at three points in time (see Figure 2 for timeline): at baseline, immediately after the end of program delivery (endline 1) and one year after the end of program delivery (endline 2). To shed light on whether parents change their language-related child investment strategies because of the intervention, we augment the data with register-based information about book loans at public libraries. We combine these information sources with population-wide register-based data maintained by Statistics Denmark. Access to the latter yields a range of advantages; it allows us to compare participating municipalities with other Danish municipalities as well as participating preschool children to other preschool children within the TipsByText municipalities. Through this, we speak to the representativeness of our study; it enables us to investigate balance in

observable characteristics across treatment and control groups; and facilitates heterogeneity analyses. Finally, we link these individual level data to the number of text messages received over time.

4.1 Measuring outcomes

We base our child outcome measures on the Danish language assessment tool. There are five different versions of the tool depending on the age of the child; each version consists of age-appropriate subtests measured with age-appropriate items. There are four versions targeting preschool children at different ages (3 years, 4 years, 5 years, and 6 years) and one version targeting schoolchildren in their first year of primary school. The older children are assessed in more dimensions of language development than that of the younger children. Specifically, the children's communication strategies are assessed at ages 3-5; their comprehension, vocabulary, and rhyme detection skills are tested at all ages; their concepts of print is tested at ages 4-6; their letter identification and deletion skills are tested at ages 5-6; and their auditory discrimination skills are tested only in primary school.[8] Furthermore, the difficulty of the questions within each subtest increases with age to avoid ceiling or floor effects. The preschool versions of the assessment are conducted one-on-one between a preschool teacher or language specialist, who is blind to the child's treatment status, and the child. For most subtests, the teacher asks the questions, the child answers, and the teacher notes whether the answer was "correct," "incorrect" or "not answered." The subtest "Communication strategies" is answered by the pedagogue/language specialist; it asks about the child's general use of language and the teacher fills out whether the child "never," "rarely," "often," or "always" use a specific communication strategy or react in a specific way. The school version of the assessment is a class-based written test: The teacher asks the questions in class and the children note their answers individually. The children who score low on the class-based test are also assessed individually using the preschool test for 6-year-old children.

To create the composite measure of language skills, we pursue a strategy that follows the spirit of York, Loeb and Doss (2019) (see also Anderson, 2008). We construct a linear index in the following way: First, for each child, we calculate the score within each subtest as the number of correct answers in that subtest. For the dimension "communication strategies," we calculate the average score on the four-point scale across the 15 questions. Second, we standardize the scores for each subtest within

[8] All language dimensions are translated from Danish as follows: Comprehension = sprogforståelse, vocabulary = ordforråd, rhyme detection = rim, concepts of print = opmærksomhed på skrift, letter identification = bogstavkendskab, deletion = opdeling af ord, and auditory discrimination = forlyd.

each test version (3 years, 4 years, 5 years, 6 years, school). Third, for each child, we average the standardized scores across subtests to get a single score. Finally, we standardize the scores within each test version again to ensure that the outcome measure has a mean of zero and a standard deviation of 1 within each test version. In creating this index, we assume that all measured language dimensions are related through a single latent (unmeasured) variable, "language skills," and that the treatment effect on the individual language dimensions operates only through this latent variable. Thus, we assume that the subtests are independent of each other conditional on the latent variable. Under this assumption, a linear index (which estimates the latent factor) "yields an adequate test by weighting each outcome by its signal-to-noise ratio" (Ludwig et al., 2017).

Absent a direct measure of parental language-related child investments, we consider their use of public libraries, while acknowledging that several other tools exist. In practice, we consider total loans as well as loans from the children's section in both 2020 and 2021.

5. Effects of TipsByText

5.1 Children's language skills

This section presents estimates of the effect of being offered the intervention on children's language skills; or intention-to-treat effects. In practice, we compare the TipsByText treatment group with the control group by running a regression of the standardized language assessment outcome on the treatment indicator controlling for the standardized language assessment at baseline (dummying out missing baseline observations), and age and preschool fixed effects. Thus, our main specification controls for a more detailed version all our randomization stratification variables.[9] We also explore whether the results are sensitive to including varying sets of controls (exact age in November 2020, and child and parental characteristics[10]). All standard errors are clustered at the family level.

Before initiating the formal analysis, we carry out a series of initial analyses to explore whether the treatment and control groups are balanced on important characteristics, and to investigate the role played by attrition from the study. We carry out our baseline balance checks by comparing the means

[9] E.g., we control for the continuous measure of language assessment whereas we stratified on a dummy for whether the child scored above or below the median. We also run versions with strata indicators exactly as they are defined in the randomization process. Results are robust to this exercise and available upon request.
[10] One specification uses the machine learning method, LASSO, to select relevant child and parental characteristics among the full set.

of an enriched set of child and parent characteristics and baseline measures of the outcome across the treatment and control groups. Table 1 performs simple comparisons of the treatment and control groups at baseline. In practice, to detect whether the groups are statistically significantly different from each other, we perform regressions of each variable on the randomization indicator while clustering standard errors at the family level and controlling for age and preschool fixed effects as in our main empirical specification. For reference, we also run versions (Table A1) that do not control for age and preschool fixed effects. As should be expected, the results indicate no obvious problems with balance; none of the mean variable-by-variable differences is large and only few are statistically significant.

Next, we investigate the role of attrition. Attrition is defined as lack of endline data for the language assessment outcome.[11] There are two types of attrition: 1) children who are still enrolled in the participating preschools but whose endline language assessment is incomplete for unspecified reasons (e.g. the children traveling during the interval for endline data collection, or the teacher being unable to get the child to answer a sufficient number of questions despite several attempts), or 2) children who moved out of the participating preschools and could not be tracked down for data collection. Importantly, a child's attrition does not imply that parents drop out of TipsByText.[12] Figure 1 shows that attrition just after program delivery is not a major concern overall; 15% of children (256 out of 1743) in the treatment group and 14% of the children (244 out of 1738) in the control group are observed with no language assessment. Importantly, families that move out of the municipality drive almost two-thirds of the missing endline measurements; in all likelihood, movement choices are entirely unrelated to their participation in TipsByText. A much smaller share of children has a successful measurement at the one-year follow-up (52% in the control group and 50% in the treatment group); the additional attrition from the first endline to the second is primarily driven by the fact that children age out of the language test after preschool (indicated as "exiting" in Figure 1), which is also entirely unrelated to their participation in TipsByText. When we exclude those who age out of the test, we arrive at attrition rates of 25% in the treatment group (904 out of the 1210 who did not age out) and 29% in the control group (865 out of the 1214 who did not age out).

Table A2 compares characteristics of children without the language assessment outcome across the treatment and control groups, and Table A3 compares characteristics of children who attritted with

[11] We define a measurement as being complete if the child answers at least 50 % of the questions.
[12] We explicitly explore drop-out in Section 6 below.

those who remained in the study sample. Overall, there are no statistically significant differences in terms of the baseline language assessment when comparing those who attrit with those who remain in the study (Table A3). There is, however, a tendency for those who attrit in the treatment group to have slightly higher language assessment scores at baseline. The differences are mostly small and not uniform across all subtests, so we do not worry that this substantially threatens internal validity. However, to address the small differences at baseline and to increase statistical power, we control for baseline language assessment scores in our main specifications.

Interestingly, the selection mechanism behind overall attrition does not appear to be uniform: first and second-generation immigrant children, children in single-parent households, and children of fathers who work part time (as compared to full time) are more likely to be missing an endline language assessment. This is in line with expectations and likely driven by the fact that language assessments are more likely to be missing for children with less well-developed language skills.

Table 2 continues to show estimated effects of TipsByText on the standardized language outcome for the overall population. Panel A shows the results immediately after program delivery; Panel B shows the results at the one-year follow-up. Column 1 controls for age and preschool fixed effects; column 2, which is our main specification that controls for all strata, adds controls for the standardized baseline measure and an indicator for whether the baseline measure was missing; column 3 is similar to column 2 but omits preschool fixed effects to include the full sample; and column 4 controls for all of these and adds characteristics measured at baseline. This last specification uses the machine learning method, the post-double selection LASSO, to select relevant child and parental characteristics among the full set of variables (Belloni et al., 2014; Ludwig et al. 2019).[13]

Contrary to initial expectations, we find no evidence that TipsByText improved children's language skills. Regardless of the conditioning set, effects are all small and precisely estimated both immediately after program delivery and at the one-year follow-up; we can reject even small positive (and negative) effects. Table 3 investigates differential effects by pre-defined subgroups for which

[13] The full set of variables included the child's sex, age, gestational age at birth, and number of siblings, father's earnings, mother's earning, father's age, mother's age, and dummies for whether the child lives is a single parent family, the father's highest education is high school or lower, the mother's highest education is high school or lower, the child lives outside of mother's and father's homes, the child is a first or second generation immigrant, the child is first born, the mother committed any crime after the birth of the child, the father committed any crime after the birth of the child, the mother had contact to psychiatric hospital after the birth of the child, the father had contact to psychiatric hospital after the birth of the child, the child had low birthweight (2,500 gram), the child had very low birthweight (1500 gram), the child had an APGAR score below 7, the child's APGAR lower than 10, and missing variable indicators for all of these variables.

we expected to see a higher treatment effect because the potential for improvement was larger: whether the child scored below the median in the baseline language assessment, whether the child's mother's or father's highest education was high school or lower, and whether the child was a first or second-generation immigrant. Thus, we find no evidence that TipsByText improved language skills in any of these pre-defined subgroups.[14]

5.1 Robustness analyses

We carry out a range of robustness analyses to explore the sensitivity of our findings to various modelling choices. We pay particular attention to functional form assumptions on the outcome. Table 4 first shows estimated effects of TipsByText on each of the (standardized) subtests of the language assessment. These include comprehension, vocabulary, rhyme detection, concepts of print, letter identification, deletion, and auditory discrimination, and communication strategies assessed by the teacher. We find no evidence that the text-messaging program improved language skills in any of these dimensions. The estimated effect on comprehension immediately after program delivery is even negative and marginally statistically significant.

Table A4 next investigates the effects of TipsByText on falling into each of the quartiles of the language assessment. This is primarily to test whether receiving the text messages might have influenced the probability of belonging to either tail of the distribution despite there not being an average effect of the program. However, we find no evidence of this—a conclusion that is confirmed by visually comparing the full distribution of endline language assessment scores between the treatment and control groups (see Figure 3). Table A5 takes a different route and instead investigates the effect of TipsByText on the share of correct answers, the share of wrong answers, the share of missing answers, and the difference between endline and baseline in the share of correct answers. Regardless of the specification, though, we are unable to detect any meaningful effects of the intervention, and all point estimates are very close to zero and precisely estimated.[15]

6. Why did TipsByText not affect children's language development?

[14] To explore heterogeneity in the effects of TipsByText beyond the pre-specified subgroups, we also employ machine learning inference on heterogenous treatment effects in line with Chernozhukov et al. (2018). We find no evidence of positive effects in any subgroups; results are available upon request.

[15] We have also explored whether results vary by whether the mother or the father received the text messages. We found no evidence of this.

Given our findings from above, it is relevant to explore channels behind the lack of effects. We conjecture that several factors related to the intervention could play a role, including delivery of the text messages, parents opting out of the program and their attitudes towards the texts more generally, and the match between the content of the texts and the child's skills. More broadly, we also explore the role of the context in explaining our results. We employ quantitative data as well as insights from the concurrent qualitative evaluation to shed light on each of these factors.

Was TipsbyText delivered as intended? Figure 4 shows the share of texts that were successfully delivered across time. Except for March 2020 where there was a technical issue with the platform that sent out the text messages, more than 90% of all parents successfully received 90-100% of all text messages. Accordingly, we are not concerned about lack of delivery as a key explanation behind the null findings.

Did families discontinue the intervention? We explored the nature and role of attrition above, defined as lack of an endline language assessment. Another relevant indicator is active drop out from the intervention. Figure 5 shows the share of families who discontinue their participation in TipsByText over time. Initially, drop out is very low but naturally, it increases over time. By the end of intervention delivery in September 2020, roughly 27.94% of parents had unsubscribed from the text-messaging program - 27.33% of mothers and 28.54% of fathers (see Appendix Figure A1).[16] A related question is whether certain types of families were more likely to discontinue the intervention – particularly those for whom we expected to see the largest effects. Figure 6 continues to demonstrate the association between children's baseline language skills and the number of messages received. Importantly, there is only a very weak relationship between the proportion of messages received and the child's baseline language skills. To sum up, it is unlikely that drop out, including selection into discontinuing the intervention, is the primary driver behind our null results.

Parents' use of the texts and the match between content and child skills. Since neither delivery nor drop out seem to have fundamental bearings for our results, we move on to explore parents' attitudes towards the texts; if parents find the information conveyed in the text messages irrelevant, for example, it is unlikely that TipsByText will positively affect child outcomes. To learn about this, we draw in the findings from the concurrent qualitative evaluation (Tolmer et al., 2021). As we mention

[16] An alternative specification that estimates the effect of being continuously subscribed to the program, while instrumenting with treatment status (i.e., treatment on the treated) yields results that are similar to our main specification. Results are available upon request.

above, 120 families, "the Interview group", were drawn at random to form the gross list for the qualitative interviews. Of these, 76 were invited to be interviewed whereof 36 declined the invitation. In the end, the team interviewed 40 parents who continued to receive the TipsByText messages throughout the trial as well as 20 parents who unsubscribed from receiving TipsByText. Overall, the qualitative evaluation found that most parents interviewed appreciated the TipsByText program and engaged with the content of the text messages. The evaluation also show that the program prompted parents to dedicate more time more regularly to their child's language development and gave parents inspiration for new or alternative language development activities to do with their children. While the methodology is different, this is in line with the findings from the existing evaluations of similar interventions (York et al., 2019; Mayer et al., 2019). A brief *ex post* parent survey carried out via text messaging confirm that parents did remember the text messages and made use of them; see Figure 7. A substantial share of parents who responded to the survey would also be willing to recommend TipsByText to others.[17]

However, while many parents were positive about the program, the qualitative evaluation also documents that roughly 25% of interviewed parents engaged somewhat less with TipsByText. Tolmer et al. (2021) characterized parents as having low or limited engagement with the program if they i) only skim read the texts or only read a few texts from time to time, and ii) did not reflect on or use the tips included in the texts. Lack of engagement was primarily either due to confusion about why they received the texts, to the parents viewing the content of the texts as 'obvious'/providing no new information, or to the parents perceiving the content as not being age appropriate for their children. Furthermore, it is possible that the share of parents who displayed low or limited engagement with the program is, in reality, considerably larger than 25%, since individuals with a more positive attitude towards the text-messaging program are probably more likely to accept an interview invitation. Finally, Tolmer et al. (2021) found that there was a tendency for engagement to decline over time. Hence, both the opt-out design and the lack of individual-level (time-varying) targeting ensuring that the content of the text messages matched the child skills might have had negative consequences for parents' attitudes towards and use of TipsByText.[18] Doss, Fahle, Loeb and York (2019), for example, show that differentiated and personalized text messages were superior to the standard version of Ready4K! Moreover, the qualitative interviews and the ex-post survey focused on parents' own

[17] Unfortunately, only 219 parents initiated the ex-post survey, and only 186 completed it.
[18] It is important to keep in mind that the opt-out design did support broad outreach; something that most likely would have been much more difficult with an opt-in design.

perceptions, and while positive attitudes towards and subsequent appropriate use of TipsByText are prerequisites for any effects on children, it does not guarantee improvements in children's language skills (though it could possibly have affected the quality of their relationship more generally if it increased the quantity and quality of time spent together).

Parental behaviors. Another way to explore whether TipsByText brought about changes in family interactions is to directly analyze behaviors related to reading. To this end, Table 5 shows the effects of TipsByText on parents' total library loans as well as loans from the children's section in both 2020 and 2021. Panel A shows the results for the entire sample and Panel B shows those for the sample with a successful first endline measure on the language assessment. One benefit of using this administrative data is that it does not suffer from the same level of attrition as does our main language outcome. However, in line with the findings on children's language attainment, we detect no effects on this measure of parental behavior.

Covid-19 and the broader context. Our evaluation coincided with the beginning of the Covid-19 pandemic. As most countries, Denmark was in a lockdown during the spring of 2020, but preschools opened before the summer of 2020 and continued to be open during the fall of 2020. It is *ex ante* theoretically unclear how the lockdown, however short relative to other countries, might have affected our findings. On the one hand, many parents were forced to be at home with their children and thus had more time to engage with them; on the other hand, Covid-19 was a substantial stressor that might easily have distracted parents and attempting to work from home while watching a child might have negatively affected the quality of parent-child interactions. Figure 7 shows that 64% of parents who participated in the *ex post* survey indicated that their use of TipsByText was unchanged by the pandemic, while 30% indicated that they used the text messages more during the lockdown.

Finally, it is possible that the population under study in York et al. (2019) is substantially worse off in terms of their socio-economic background and associated child language development when compared to our Danish equivalent; and it is likely that the gains from an intervention like TipsByText decrease with the quality of the family environment or the context more broadly. In addition, we speculate that the quality of preschool care is lower in their context; see Esping-Andersen et al.,(2012). To learn about the socio-economic background of our study sample, we first compare families with preschool children within TipsByText municipalities to their counterparts in other municipalities. In practice, we perform regressions of each variable on a TipsByText municipality indicator while clustering at the family level. Table 6 shows that participating municipalities are

somewhat – but not very – disadvantaged: on average, the share of immigrants (who, for good reasons often perform worse in terms of Danish language development)[19] is 4 percentage points higher than in nonparticipating municipalities, which should be seen relative to a mean of 16%; mothers and fathers are 7 percentage points more likely to have attained at most a high school degree than those in nonparticipating municipalities; mothers earn about DKK 16.000 less annually; and fathers earn about DKK 27.000 less. This corresponds to mothers earning 6% less compared to her counterparts in other municipalities (8% of a standard deviation) and fathers earning 7% less (8% of a standard deviation) less. Thus, the participating municipalities are somewhat worse off compared to non-participating municipalities, but the differences are small, and the median family in the participating municipalities is not disadvantaged in absolute terms: The average household income is roughly $90,000, and 47.55% of mothers have some post-secondary education. Table 7 continues to compare our study population to other families with preschool children within the TipsByText municipalities. Again, there is indication of some disadvantage but not a lot; the share of immigrants if 4 percentage points higher; mothers (fathers) are 9 (10) percentage points more likely to have at most a high school degree; mothers earn roughly DKK 13,000 (5%) less and fathers about DKK 25,000 (6%) less.

We conclude that the Covid-19 pandemic probably did not affect our findings much but that the general context may have. York et al. (2019) targeted the poorest third of San Francisco's preschool market, whereas our study population is only slightly disadvantaged when compared to the overall Danish population, and is, on average, not disadvantaged in absolute terms; the average household income in our study population is, for example, roughly $90,000. To put this into perspective, 24% of households in the San Francisco Unified School District earn less than $50,000 and 18% earn between $50,000 and $100,000 (together, roughly the poorest third). The San Francisco Unified School District is also much more linguistically diverse; among children aged 5-17, only 56% live in English-only households.[20] In our study, only about 20% are first- or second-generation immigrants.

7. Conclusion

This paper exploits a large randomized controlled trial that combines survey and register-based data to investigate the effects of a texting program, TipsByText. TipsByText focuses on child language

[19] In our sample, immigrants make up 30% of the children with below median language development and 12% of the group who scores above the median.
[20] https://censusreporter.org/profiles/97000US0634410-san-francisco-unified-school-district-ca/

development and aims to support parents and caregivers in creating playful, language stimulating activities. The program was delivered to parents of children aged 3-6 in Danish preschools via an opt-out design. Because it may require a longer period to alter behaviors within the family leading to improved language acquisition among children, it was likely that any impact of TipsByText on child language skills would take time to materialize. Accordingly, we measured children's outcomes both immediately after the end of program delivery and at a one-year follow up. The opt-out design enabled broad outreach but though the program was, in line with the findings from other work, well-liked by parents and delivered as intended with relatively low dropout, it did not affect children's language development.

Our results on child language development are somewhat in contrast to existing studies of a similar intervention (York et al., 2019). We conjecture that the opt-out design – because it enrolls families that are not necessarily highly engaged with the intervention – as well as the lack of individual-level (time-varying) targeting might explain the lack of effects of TipsByText. It is also possible that the general context has played a role; children in our study, while somewhat disadvantaged compared to the overall population, are likely to have grown up in homes with more resources than those in the US studies, just as the quality of nonparental care is likely to be higher.

Literature

Almond, D., J. Currie and V. Duque (2018): "Childhood Circumstances and Adult Outcomes: Act II," the Journal of Economic Literature 56(4), 1360-1446.

Andersen, S. C. and H. S. Nielsen (2016): "Reading intervention with a growth mindset approach improves children's skills", PNAS 113 (43), 12111-12113.

Anderson, M. L. (2008): "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." Journal of the American statistical Association 103, 1481–1495.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). "High-Dimensional Methods and Inference on Structural and Treatment Effects." Journal of Economic Perspectives, 28 (2): 29-50.

Bleses, D., W. Vach, R. N. Jørgensen and T. Worm (2010): "The internal validity and acceptability of the Danish SI-3: A language-screening instrument for 3-yearolds." Journal of Speech, Language, and Hearing Research 53, 490-507.

Carneiro, P. and J. J. Heckman (2003): "Human Capital Policy", in J. Heckman and A. Krueger, eds., Inequality in America: What Role for Human Capital Policy?, (MIT Press).

Chernozhukov, V., I. Fernández-Val and Y. Luo (2018): "The sorted effects method: discovering heterogeneous effects beyond their averages". Econometrica 86(6), 1911-1938.

Doss, C., E. M. Fahle, S. Loeb and B. N. York (2019): "More Than Just a Nudge: Supporting Kindergarten Parents with Differentiated and Personalized Text Messages", Journal of Human Resources 54(3), 567-603.

Dweck, C. S. (2007): "Mindset: The new psychology of success". Random House Digital, Inc.

Esping-Andersen G., I. Garfinkel, W. J. Han, K. Magnuson, S. Wagner and J. Waldfogel (2012): "Child Care and School Performance in Denmark and the United States", Child Youth Services Review 34(3), 576-589.

Heidlage, J. K., J. E. Cunningham, A. P. Kaiser, C. M. Trivette, E. E. Barton, J. R. Frey and M. Y. Roberts (2018): "The effects of parent-implemented language interventions on child linguistic outcomes: A meta-analysis." Early Childhood Research Quarterly 50, 6–23.

Ludwig, J., S. Mullainathan, and J. Spiess (2019): "Augmenting pre-analysis plans with machine learning." In AEA Papers and Proceedings 109, 71–76.

Mayer, S., A. Kalil, P. Oreopoulos and S. Gallegos (2019): Using behavioral insights to increase parental engagement: the parents and children together intervention. Journal of Human Resources 54(4), 900-925.

Ministry of Children and Education (2019): Guide to the Language Assessment for 3-6 Year Olds ("Vejledning til Sprogvurdering 3-6).

Nielsen, H., P. Jensen, D. Bleses, J. Ostersen, L. Dybdal and S. Breining (2017): "Fremtidens dagtilbud",

Reese, E., A. Sparks and D. Leyva (2010): "A Review of Parent Interventions for Preschool Children's Language and Emergent Literacy." Journal of Early Childhood Literacy 10(1), 97–117.

Riis-Vestergaard, M. (2021): *TipsByText: A Parent-Oriented Language Intervention for Preschool Children*. Unpublished doctoral dissertation, Princeton University, Princeton. http://arks.princeton.edu/ark:/88435/dsp01pv63g3376

OECD (2006): "Starting Strong II - Early childhood education and care." Technical report.

Rockwool Foundation (2019): "Adaption of the US TipsByText Curriculum to a Danish context".

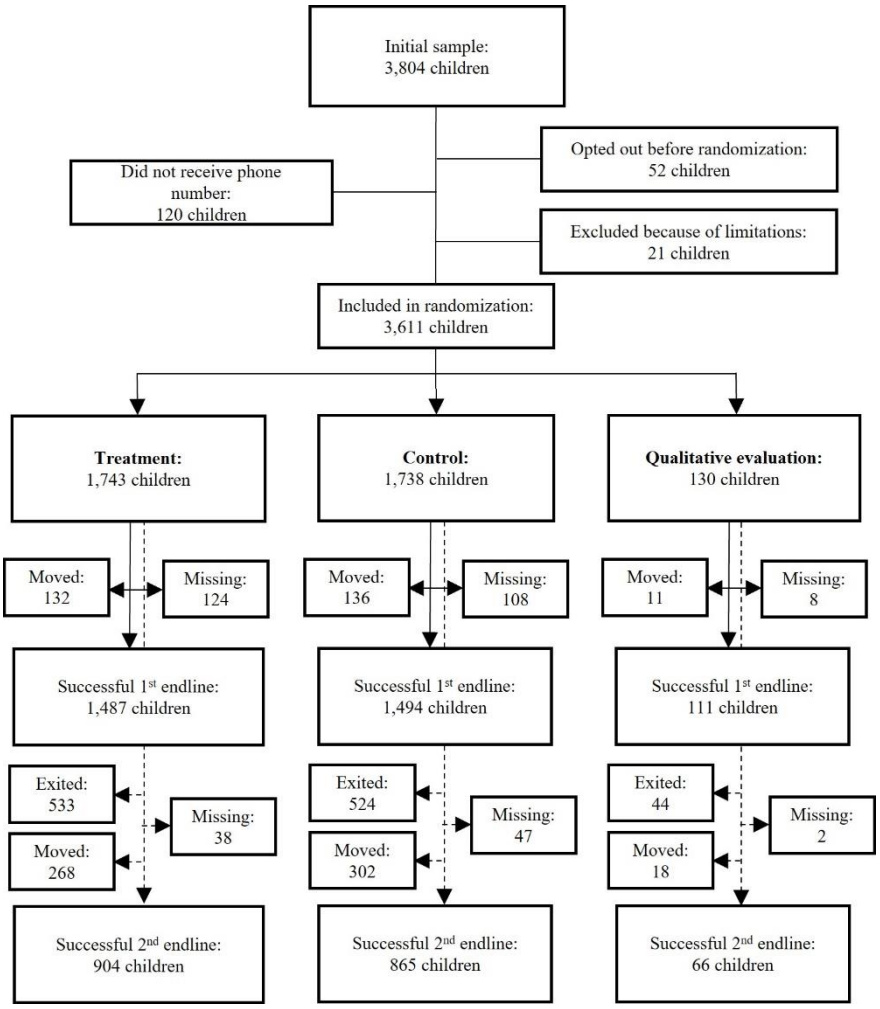
Tolmer, Z. (2021): "Qualitative evaluation report of TipsByText in Denmark", CoInt report.

York, B. N., S. Loeb and C. Doss (2019): One Step at a Time: The Effects of an Early Literacy Text-Messaging Program for Parents of Preschoolers, Journal of Human Resources 54(3), 537-566.

Tables and Figures

Figure 1

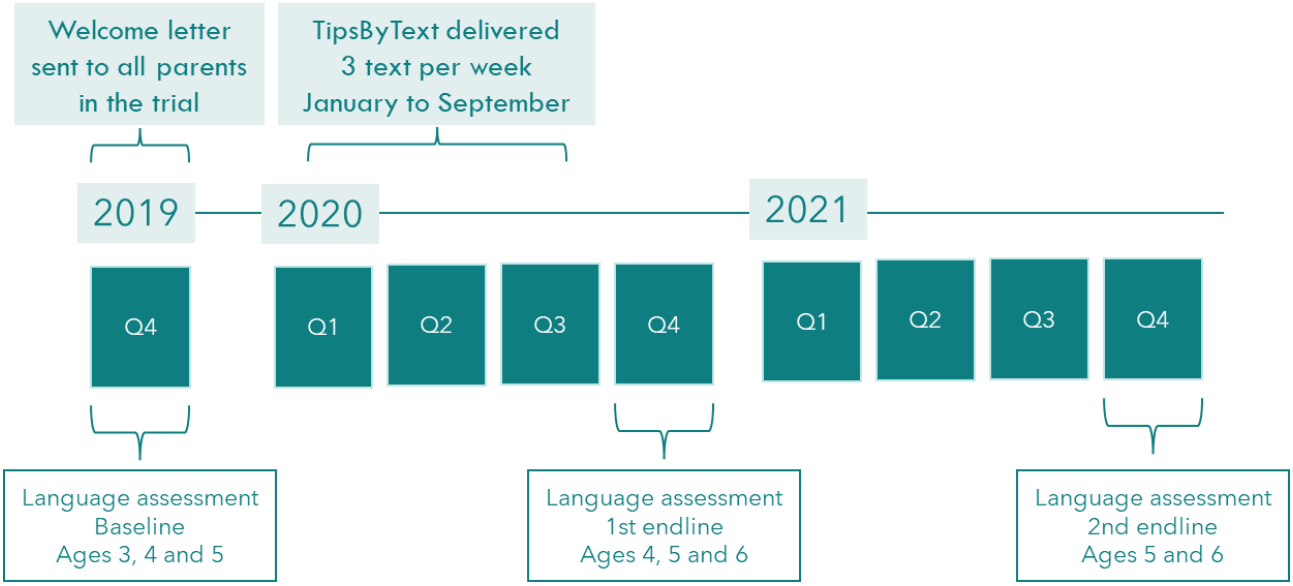
TipsByText Flowchart



Notes: Figure presents the research design as well as the final sample size. 3,804 children were enrolled in participating preschools. To be eligible for treatment, a phone number was required, which resulted in the exclusion of 120 children. Parents of 52 children chose to opt-out, when they received the information material. Furthermore 21 children were excluded due to limitations. This meant 3,611 children were included in the randomization. 1,743 children were assigned treatment, 1,738 were assigned to be control group and 130 children were randomized to receive treatment and possibly partake in the qualitative evaluation. Of the 1,743 children who received treatment, some moved out of the municipality or to private non-participating schools when they turn 6 (132) and some (124) didn't have a successful 1st endline language assessment. For the second endline language assessment, children could withdraw from the evaluation for one of the reasons above, or they could exit the evaluation by turning 7 and therefore are outside the age range of conducting a suitable language assessment.

Figure 2

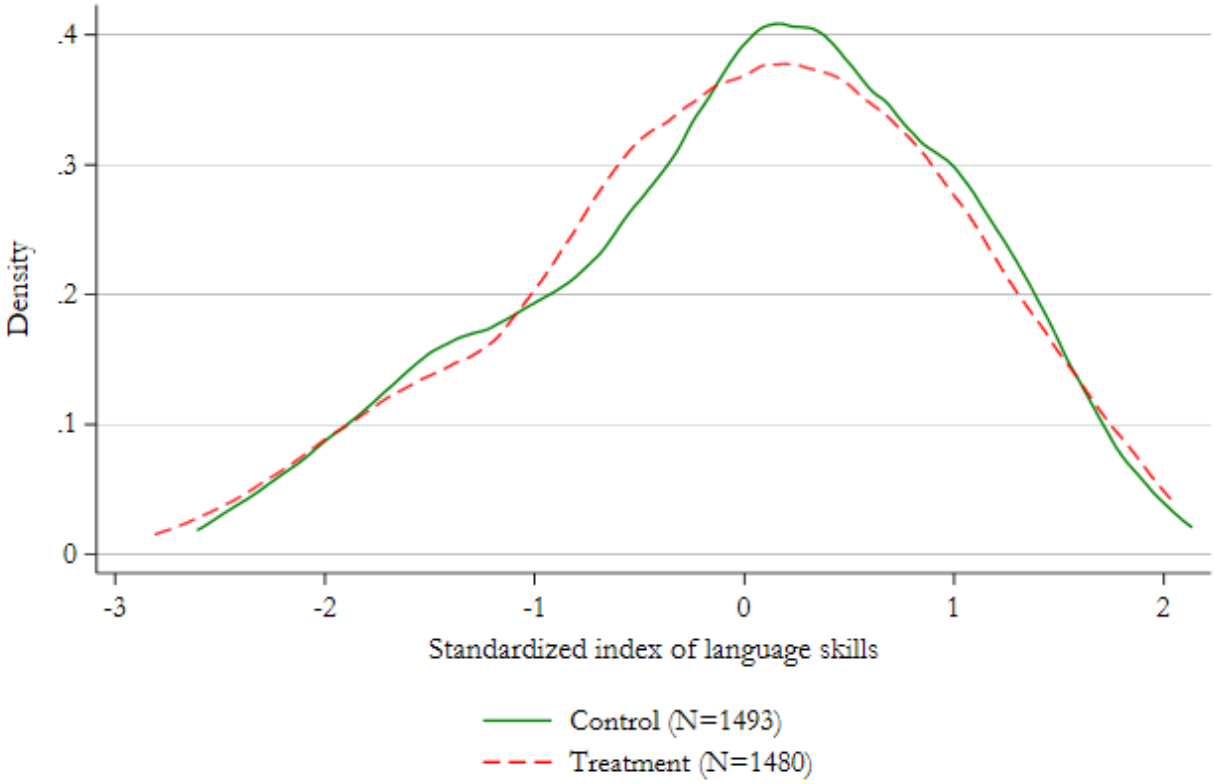
Timeline for the evaluation of TipsByText in Denmark



Notes: Figure presents the timeline of the intervention. In Q4 of 2019, parents received a welcome letter giving detailed information regarding the trial and their options for opting-out if they chose to. After the welcome letter was distributed, the baseline language assessment was conducted. In January 2021 the intervention began, and text messages was sent to parents over the course of eight months. Language assessments were done primarily in October and November 2020 1st endline language assessment was conducted. A year later in October and November 2021, the second endline language assessments was done.

Figure 3

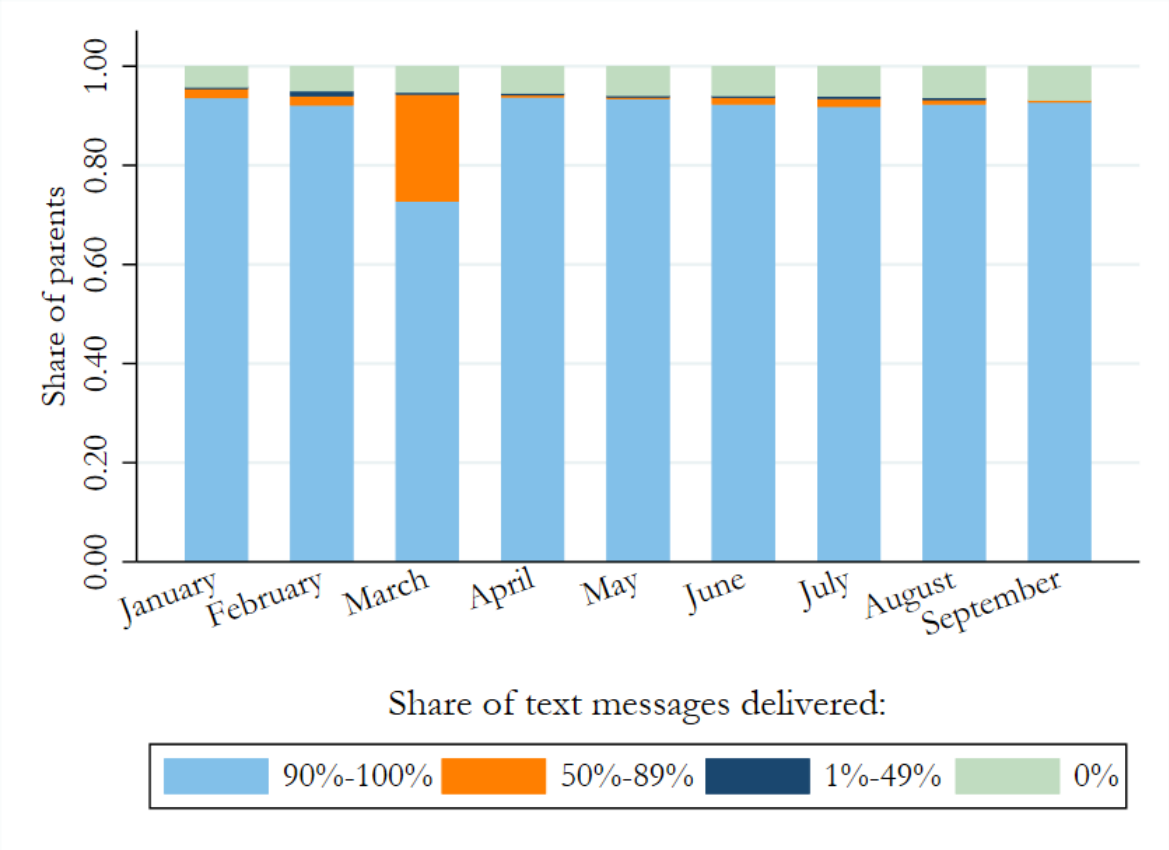
Distribution of language skills at 1st endline measurement in the treatment and control group



Notes: The figure present density plots of the outcome distribution of language skills in the treatment group (red/dashed) compared to the control group (green/solid). The density plot is constructed with an Epanechnikov kernel function.

Figure 4

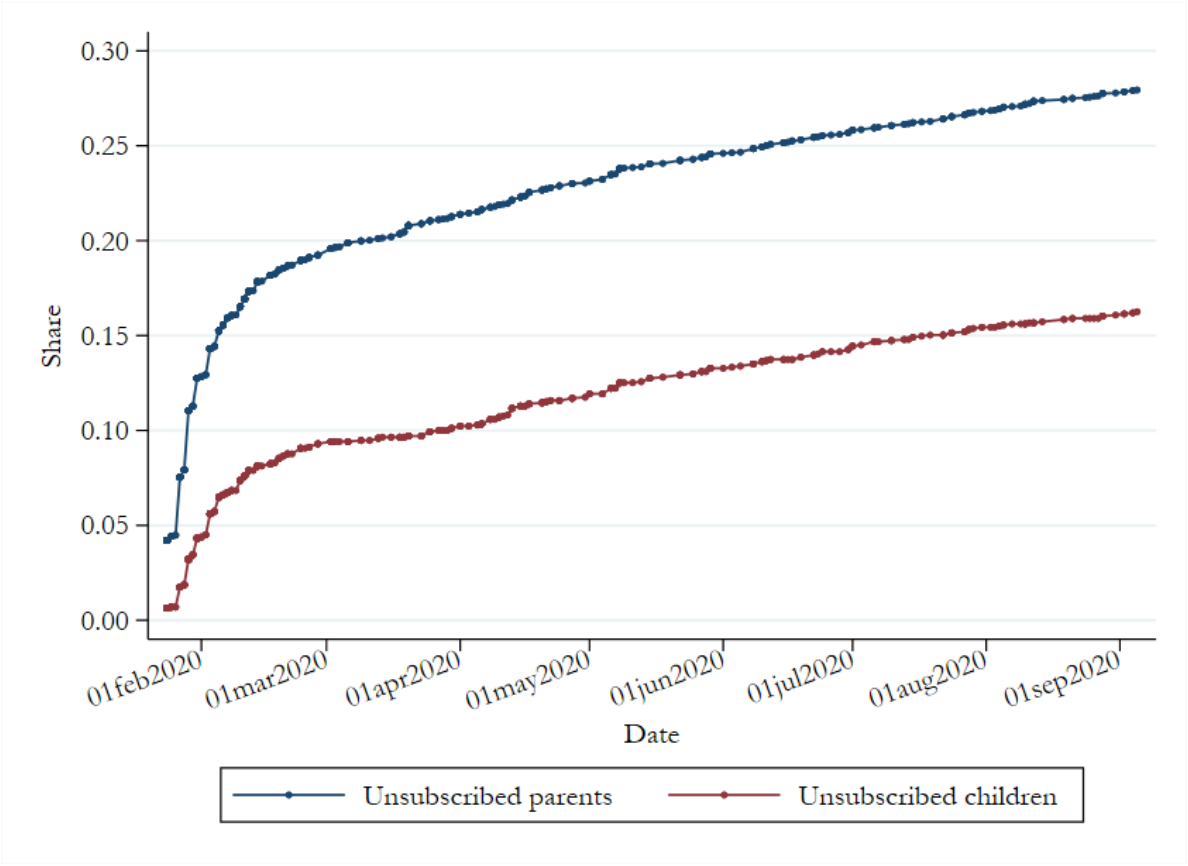
Text messages received



Notes: This figure shows the share of successfully delivered text messages by month.

Figure 5

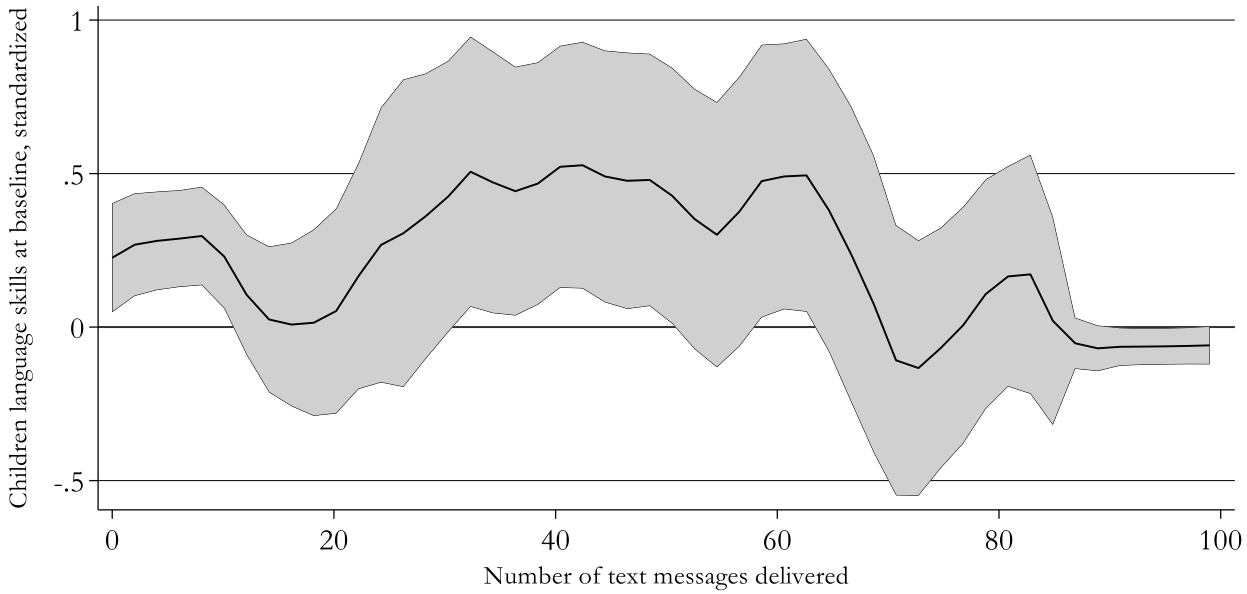
Unsubscribed parents and children



Notes: This figure shows the share of parents who unsubscribed from the TipsByText program (blue line), and the share of children for whom both parents unsubscribed from the program (red line).

Figure 6

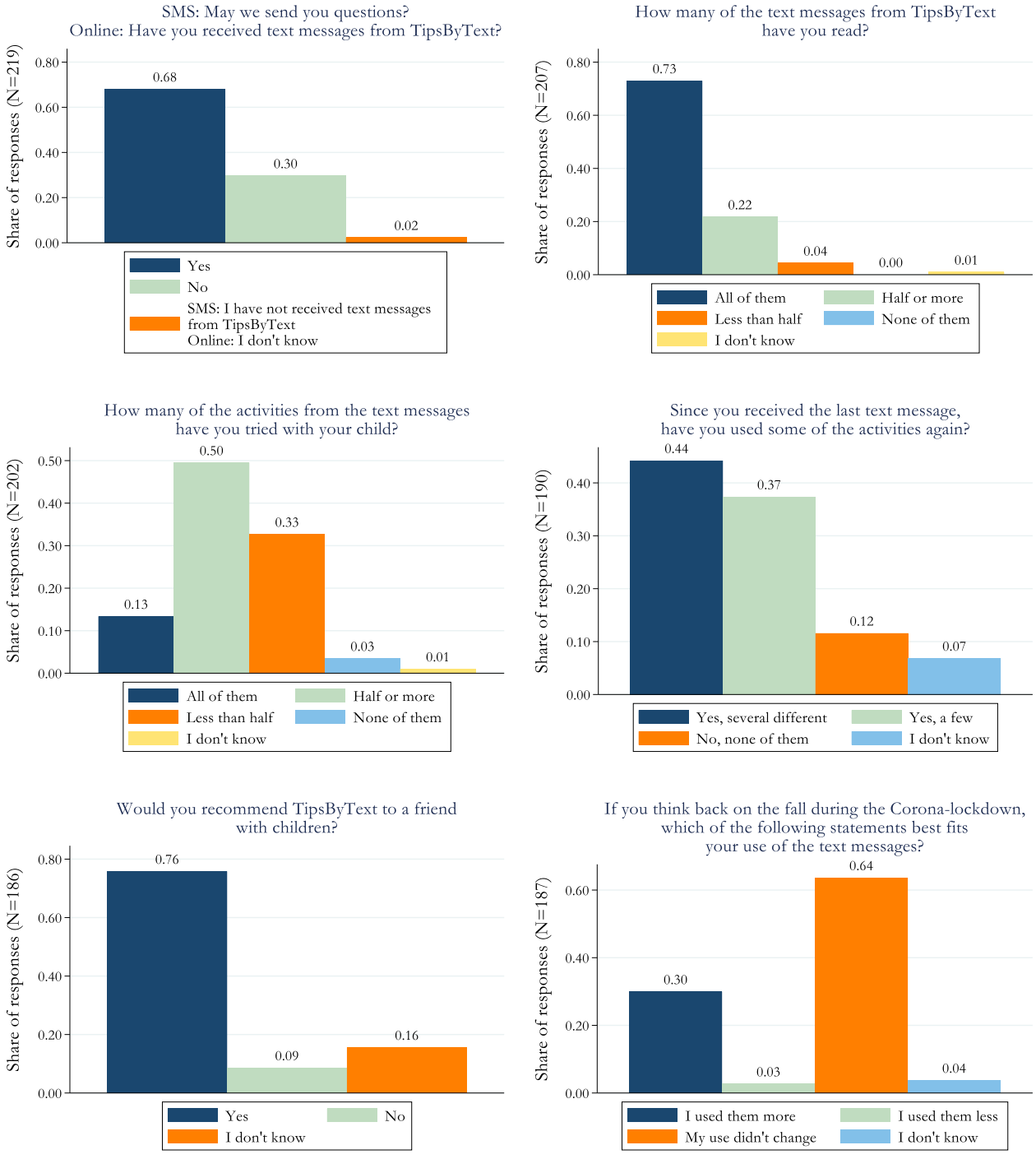
Child language skills at baseline, by number of text messages received



Notes: This figure shows children's language skills at baseline by the total number of text messages delivered.

Figure 7

Ex post survey



Notes: This figure presents the results of the parent survey conducted via SMS at the end of the program. Each panel shows the share of parents who responded in a certain way to the question indicated above that panel.

Table 1

Randomization checks: Comparing the control and treatment groups at baseline

	Control mean (SD)	Difference between treated and control group (SE)	N Treated	N Control
Stratification variables				
Indicator: Below median at baseline test	0.473	-0.010	1487	1494
	(0.499)	(0.018)		
Indicator: Estimated endline test version = 3	0.033	0.000	1487	1494
	(0.178)	(0.004)		
Indicator: Estimated endline test version = 4	0.345	-0.010	1487	1494
	(0.476)	(0.006)		
Indicator: Estimated endline test version = 5	0.345	0.010	1487	1494
	(0.476)	(0.008)		
Indicator: Estimated endline test version = School (group)	0.284	0.000	1487	1494
	(0.451)	(0.007)		
Baseline language assessment variables				
Baseline: Total % correct	0.477	0.000	1385	1401
	(0.200)	(0.007)		
Baseline: Total % wrong	0.424	0.010	1385	1401
	(0.173)	(0.006)		
Baseline: Total % missing	0.099	0.000	1385	1401
	(0.153)	(0.005)		
Baseline: Comprehension, % correct	0.656	0.000	1385	1401
	(0.190)	(0.007)		
Baseline: Vocabulary, % correct	0.417	-0.010	1385	1401
	(0.238)	(0.008)		
Baseline: Rhyme detection, % correct	0.472	0.000	1385	1401
	(0.356)	(0.011)		
Baseline: Concepts of print, % correct	0.473	0.000	852	901
	(0.261)	(0.011)		
Baseline: Letter identification, % correct	0.377	-0.040*	403	416
	(0.312)	(0.022)		
Baseline: Deletion, % correct	0.450	-0.010	403	416
	(0.297)	(0.022)		
Baseline: Communication, mean	2.752	-0.010	1365	1376
	(0.657)	(0.022)		
Baseline: Age (years) at language test	4.329	-0.040***	1385	1400
	(0.872)	(0.013)		

Notes: Table shows means for the control families in the TipsByText project (first column) and results from regressions of characteristics on the randomization indicator, controlling for test version and preschool fixed effects (second column). Stars indicate significance levels from regressions using cluster-robust standard errors on the family level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1 continued
Randomization checks: Comparing the control and treatment groups at baseline

	Control mean (SD)	Difference between treated and control group (SE)	N Treated	N Control
Demographic variables				
Indicator: Boy	0.497	0.030	1487	1494
	(0.500)	(0.019)		
APGAR-score < 10 (max)	0.070	-0.010	1425	1433
	(0.255)	(0.009)		
Low birth weight (< 2500 g)	0.054	0.000	1425	1433
	(0.227)	(0.009)		
Number of siblings in intervention	1.292	0.010	1479	1484
	(1.092)	(0.042)		
Indicator: Firstborn child (mother)	0.442	-0.020	1487	1494
	(0.497)	(0.018)		
Indicator: One adult in home	0.123	-0.020*	1479	1484
	(0.328)	(0.012)		
Indicator: First/second generation immigrant	0.194	0.000	1487	1494
	(0.396)	(0.014)		
Mother is working part time	0.184	-0.010	1477	1475
	(0.387)	(0.015)		
Father is working part time	0.078	0.000	1445	1449
	(0.268)	(0.011)		
Indicator: Dad's education high school or less	0.637	0.020	1487	1494
	(0.481)	(0.018)		
Indicator: Mom's education high school or less	0.537	0.000	1487	1494
	(0.499)	(0.019)		
Father income (thousands)	362.883	7.310	1447	1451
	(225.597)	(8.395)		
Mother income (thousands)	224.914	0.510	1477	1480
	(173.017)	(6.504)		
Father's age at birth of child	33.251	-0.250	1445	1450
	(6.002)	(0.239)		
Mother's age at birth of child	30.248	0.010	1450	1450
	(4.931)	(0.198)		

Notes: Table shows means for the control families in the TipsByText project (first column) and results from regressions of characteristics on the randomization indicator, controlling for test version and preschool fixed effects (second column). Stars indicate significance levels from regressions using cluster-robust standard errors on the family level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2

Effects of TipsByText on child language skills

	Dependent variable: Standardized index of language skills			
	(1)	(2)	(3)	(4)
Treatment	-0.041	-0.029	-0.030	-0.032
	(0.035)	(0.026)	(0.027)	(0.026)
Preschool FE	✓	✓		✓
Test version FE	✓	✓	✓	✓
Baseline controls		✓	✓	✓
LASSO controls				✓
SE cluster level	Family	Family	Family	Family
Observations	2977	2977	2981	2977
R^2	0.206	0.554	0.480	0.562

Panel A: Outcomes measured immediately after program delivery

	Dependent variable: Standardized index of language skills			
	(1)	(2)	(3)	(4)
Treatment	-0.012	0.000	0.008	0.002
	(0.043)	(0.035)	(0.039)	(0.035)
Preschool FE	✓	✓		✓
Test version FE	✓	✓	✓	✓
Baseline controls		✓	✓	✓
LASSO controls				✓
SE cluster level	Family	Family	Family	Family
Observations	1763	1763	1769	1763
R^2	0.247	0.486	0.324	0.493

Panel B: Outcomes measured at the one-year follow-up

Notes: Table presents a regression of the standardized language outcome on the treatment indicator with varying sets of controls indicated below the table. Baseline controls include the language skills measure and an indicator for whether the language skills measure was missing. Singletons in fixed effects are dropped. Stars indicate significance levels from regressions using cluster-robust standard errors on the family level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3

Effects of TipsByText, by predefined subgroups

	Dependent variable: Standardized index of language skills			
	(1) Below median at baseline	(2) Mother's education is high school or less	(3) Father's education is high school or less	(4) First/second generation immigrant
Treatment	0.016	-0.056	-0.012	0.012
	(0.039)	(0.036)	(0.033)	(0.064)
Preschool FE	✓	✓	✓	✓
Test version FE	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓
SE cluster level	Family	Family	Family	Family
Observations	1390	1598	1927	569
R^2	0.445	0.566	0.547	0.560

Panel A: Outcomes measured immediately after program delivery

	Dependent variable: Standardized index of language skills			
	(1) Below median at baseline	(2) Mother's education is high school or less	(3) Father's education is high school or less	(4) First/second generation immigrant
Treatment	0.013	-0.038	-0.006	0.032
	(0.055)	(0.051)	(0.045)	(0.088)
Preschool FE	✓	✓	✓	✓
Test version FE	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓
SE cluster level	Family	Family	Family	Family
Observations	790	932	1142	306
R^2	0.421	0.490	0.475	0.589

Panel B: Outcomes measured at the one-year follow-up

Notes: Table presents regressions of the standardized language outcome on the treatment indicator across predefined subgroups. All regressions control for age and preschool fixed effects, the standardized baseline measure, and an indicator for whether the baseline language skills measure was missing. Singletons in fixed effects are dropped. Stars indicate significance levels from regressions using cluster-robust standard errors on the family level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4

Effects of TipsByText on subscales of language assessment

	(1) Comprehension	(2) Vocabulary	(3) Rhyme detection	(4) Concepts of print
Treatment	-0.054*	0.007	0.005	-0.009
	(0.030)	(0.028)	(0.031)	(0.032)
Preschool FE	✓	✓	✓	✓
Test version FE	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓
SE cluster level	Family	Family	Family	Family
Observations	2977	2316	2977	2211
R^2	0.411	0.587	0.347	0.422
	(5) Letter identification	(6) Deletion	(7) Auditory discrimination	(8) Communication
Treatment	-0.022	-0.002	-0.108*	0.024
	(0.042)	(0.048)	(0.064)	(0.034)
Preschool FE	✓	✓	✓	✓
Test version FE	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓
SE cluster level	Family	Family	Family	Family
Observations	1830	1169	659	2031
R^2	0.323	0.399	0.391	0.435

Panel A: Outcomes measured immediately after program delivery

Notes: Table presents regressions of the standardized outcome on the treatment indicator, preschool fixed effects and test version, the standardized baseline measure, and an indicator for whether baseline was missing. Singletons in fixed effects are dropped. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4 continued

Effects of TipsByText on subscales of language assessment

	(1) Comprehension	(2) Vocabulary	(3) Rhyme detection	(4) Concepts of print
Treatment	-0.014	-0.003	-0.015	0.052
	(0.039)	(0.040)	(0.040)	(0.046)
Preschool FE	✓	✓	✓	✓
Test version FE	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓
SE cluster level	Family	Family	Family	Family
Observations	1763	1204	1763	1204
R^2	0.390	0.567	0.355	0.423
	(5) Letter identification	(6) Deletion	(7) Auditory discrimination	(8) Communication
Treatment	-0.008	0.050	0.009	0.012
	(0.046)	(0.051)	(0.070)	(0.055)
Preschool FE	✓	✓	✓	✓
Test version FE	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓
SE cluster level	Family	Family	Family	Family
Observations	1675	1116	558	958
R^2	0.238	0.339	0.335	0.407

Panel B: Outcomes measured at the one-year follow-up

Notes: Table presents regressions of the standardized outcome on the treatment indicator, preschool fixed effects and test version , the standardized baseline measure, and an indicator for whether baseline was missing. Singletons in fixed effects are dropped. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5

Effects of TipsByText on Parents' library loans

	(1) Loans in 2020, total	(2) Loans in 2020, children section	(3) Loans in 2021, total	(4) Loans in 2021, children section
Treatment	0.011	-0.001	-0.022	-0.029
	(0.036)	(0.037)	(0.037)	(0.037)
Age of child FE	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓
SE cluster level	Family	Family	Family	Family
Observations	3453	3453	3453	3453
R^2	0.024	0.024	0.027	0.026

Panel A: Full sample

	(1) Loans in 2020, total	(2) Loans in 2020, children section	(3) Loans in 2021, total	(4) Loans in 2021, children section
Treatment	-0.009	-0.021	-0.007	-0.008
	(0.038)	(0.039)	(0.048)	(0.048)
Age of child FE	✓	✓	✓	✓
Preschool FE	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓
SE cluster level	Family	Family	Family	Family
Observations	2959	2959	1750	1750
R^2	0.086	0.082	0.128	0.127

Panel B: Sample with successful first endline

Notes: Table presents regressions of books loans (standardized by age of the child on November 1, 2020) on the treatment indicator, controlling for child age on November 1, 2020, the standardized baseline measure, and an indicator for whether baseline was missing. Panel B adds controls for preschool fixed effects, which restricts the sample to the children who successfully completed the endline language test from where we have the preschool data. In both panels, 28 observations are missing due to missing data on loans. Singletons in fixed effects are dropped. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6

Comparing families with children in preschool in participating municipalities to those in non-participating municipalities

	Non-participating municipalities mean (SD)	Difference between participating and non-participating municipalities (SE)	N Participating municipalities	N Non-participating municipalities
Demographic variables				
Indicator: Boy	0.514	0.000	5491	178,985
	(0.500)	(0.007)		
APGAR-score < 10 (max)	0.067	0.000	5213	170,741
	(0.250)	(0.004)		
Low birth weight (< 2500 g)	0.052	0.000	5213	170,741
	(0.222)	(0.003)		
Number of siblings in intervention	1.256	0.010	5491	178,940
	(1.078)	(0.016)		
Indicator: Firstborn child (mother)	0.462	-0.020**	5491	178,985
	(0.499)	(0.007)		
Indicator: One adult in home	0.137	-0.020***	5491	178,985
	(0.344)	(0.005)		
Indicator: First/second generation immigrant	0.156	0.040***	5491	178,985
	(0.363)	(0.006)		
Mother is working part time	0.174	0.000	5374	175,087
	(0.379)	(0.006)		
Father is working part time	0.080	0.010	5295	171,682
	(0.272)	(0.004)		
Indicator: Dad's education high school or less	0.531	0.070***	5491	178,985
	(0.499)	(0.007)		
Indicator: Mom's education high school or less	0.425	0.070***	5491	178,985
	(0.494)	(0.007)		
Father income (thousands)	400.444	-27.420***	5357	173,914
	(331.871)	(3.491)		
Mother income (thousands)	247.303	-16.430***	5463	177,988
	(207.684)	(2.689)		
Father's age at birth of child	33.426	-0.180**	5339	173,259
	(6.043)	(0.090)		
Mother's age at birth of child	30.802	-0.380***	5339	173,259
	(5.005)	(0.074)		

Notes: Table shows means for the non-study municipalities and compares to municipalities participating in the TipsByText project. Second column shows results from regression of characteristic on study municipality indicator. Singletons in fixed effects are dropped. Stars indicate significance levels from associated t-test using cluster-robust standard errors on family level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7
Comparing the study population to other families with children in preschool within the participating municipalities

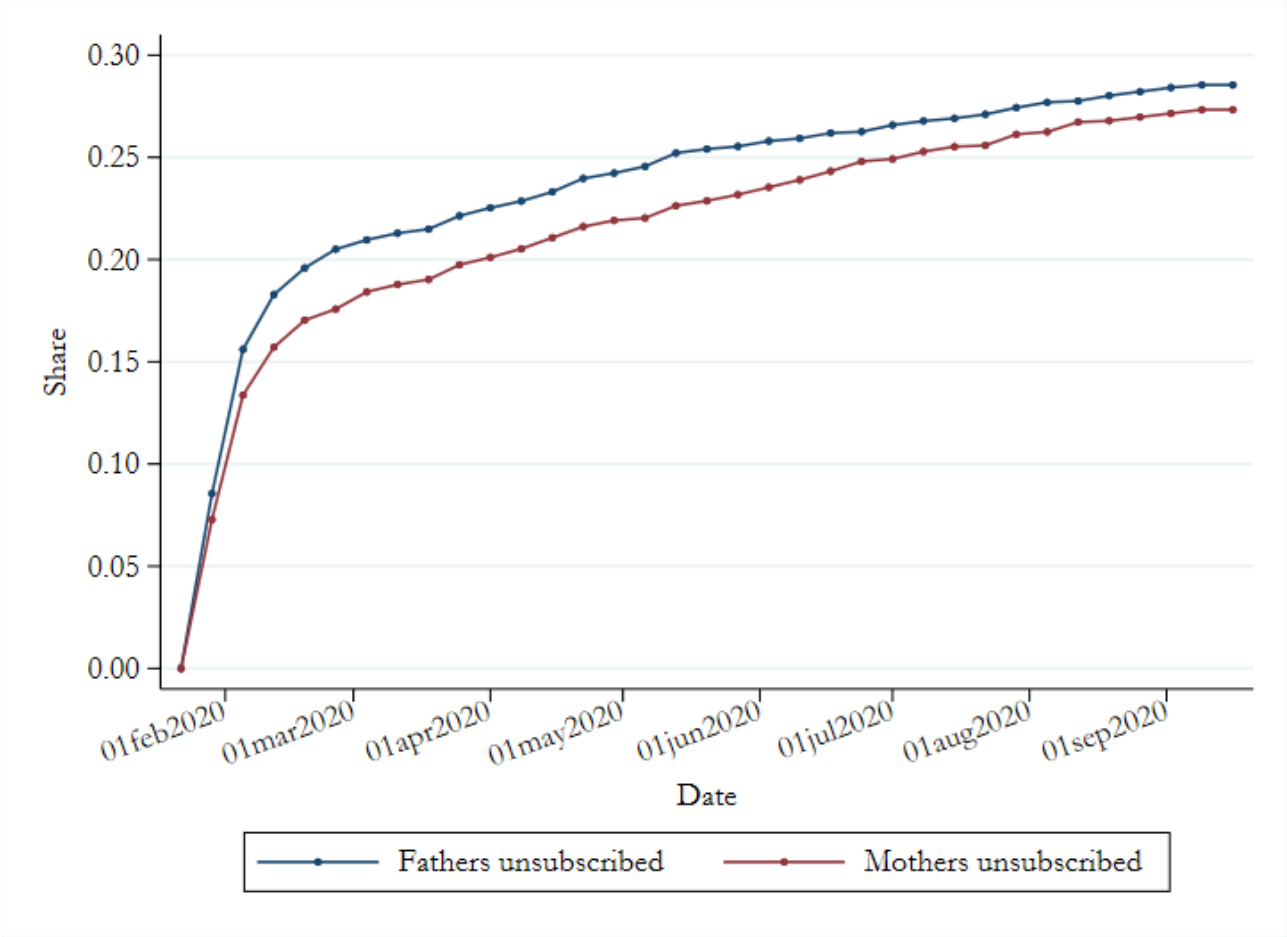
	Non-participating families mean (SD)	Difference between participating and non-participating families (SE)	N Participating families	N Non-participating families
Demographic variables				
Indicator: Boy	0.517	-0.010	3493	1998
	(0.500)	(0.014)		
APGAR-score < 10 (max)	0.076	-0.010	3348	1865
	(0.264)	(0.007)		
Low birth weight (< 2500 g)	0.047	0.010	3348	1865
	(0.212)	(0.007)		
Number of siblings in intervention	1.238	0.040	3493	1998
	(1.010)	(0.032)		
Indicator: Firstborn child (mother)	0.456	-0.020	3493	1998
	(0.498)	(0.013)		
Indicator: One adult in home	0.115	0.000	3493	1998
	(0.319)	(0.010)		
Indicator: First/second generation immigrant	0.175	0.040***	3493	1998
	(0.380)	(0.012)		
Mother is working part time	0.170	0.010	3473	1901
	(0.376)	(0.012)		
Father is working part time	0.090	0.000	3407	1888
	(0.286)	(0.009)		
Indicator: Dad's education high school or less	0.543	0.100***	3493	1998
	(0.498)	(0.015)		
Indicator: Mom's education high school or less	0.435	0.090***	3493	1998
	(0.496)	(0.015)		
Father income (thousands)	388.665	-24.540***	3415	1942
	(246.702)	(7.177)		
Mother income (thousands)	239.196	-13.060**	3482	1981
	(188.299)	(5.544)		
Father's age at birth of child	33.531	-0.460**	3403	1936
	(5.947)	(0.183)		
Mother's age at birth of child	30.789	-0.580***	3403	1936
	(4.904)	(0.149)		

Notes: Table shows means for families outside of the study population within the municipalities participating in the TipsByText project and compares to families within the study population. 'Study difference' shows results from regression of characteristic on study participation indicator. Singletons in fixed effects are dropped. Stars indicate significance levels from associated t-test using cluster-robust standard errors on family level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix tables and figures

Figure A1

Gender of unsubscribed parents



Notes: This figure shows the share of parents across gender for whom have unsubscribed from the TipsByText program. The share of fathers who have unsubscribed (blue line) is larger than the share of mothers (red line) across the period.

Table A1

Randomization checks: Comparing the control and treatment groups at baseline

	Control mean (SD)	Difference between treated and control group (SE)	N Treated	N Control
Stratification variables				
Indicator: Below median at baseline test	0.407	-0.010	1743	1738
	(0.491)	(0.017)		
Indicator: Estimated endline test version = 3	0.032	0.000	1738	1736
	(0.177)	(0.006)		
Indicator: Estimated endline test version = 4	0.328	0.010	1738	1736
	(0.470)	(0.016)		
Indicator: Estimated endline test version = 5	0.322	-0.010	1738	1736
	(0.467)	(0.016)		
Indicator: Estimated endline test version = School (group)	0.317	0.000	1738	1736
	(0.466)	(0.015)		
Baseline language assessment variables				
Baseline: Total % correct	0.479	0.000	1602	1607
	(0.202)	(0.007)		
Baseline: Total % wrong	0.423	0.000	1602	1607
	(0.172)	(0.006)		
Baseline: Total % missing	0.097	0.000	1602	1607
	(0.154)	(0.006)		
Baseline: Comprehension, % correct	0.655	0.000	1602	1607
	(0.191)	(0.007)		
Baseline: Vocabulary, % correct	0.417	0.000	1602	1607
	(0.239)	(0.009)		
Baseline: Rhyme detection, % correct	0.485	0.000	1602	1607
	(0.356)	(0.013)		
Baseline: Concepts of print, % correct	0.481	0.010	1020	1067
	(0.266)	(0.012)		
Baseline: Letter identification, % correct	0.377	-0.010	538	536
	(0.309)	(0.019)		
Baseline: Deletion, % correct	0.451	0.010	538	536
	(0.301)	(0.019)		
Baseline: Communication, mean	2.764	0.000	1573	1577
	(0.663)	(0.024)		
Baseline: Age (years) at language test	4.383	-0.030	1602	1607
	(0.876)	(0.030)		

Notes: Table shows means for the control families in the TipsByText project. Second column shows results from regressions of characteristics on the randomization indicator. Stars indicate significance levels from associated t-test using cluster-robust standard errors on family level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A1 continued
Randomization checks: Comparing the control and treatment groups at baseline

	Control mean (SD)	Difference between treated and control group (SE)	N Treated	N Control
Demographic variables				
Indicator: Boy	0.493	0.040**	1743	1738
	(0.500)	(0.017)		
APGAR-score < 10 (max)	0.068	-0.010	1654	1654
	(0.252)	(0.009)		
Low birth weight (< 2500 g)	0.056	0.000	1654	1654
	(0.230)	(0.009)		
Number of siblings in intervention	1.297	-0.020	1728	1725
	(1.109)	(0.040)		
Indicator: Firstborn child (mother)	0.447	-0.010	1743	1738
	(0.497)	(0.016)		
Indicator: One adult in home	0.127	-0.010	1728	1725
	(0.333)	(0.012)		
Indicator: First/second generation immigrant	0.207	0.000	1743	1738
	(0.405)	(0.015)		
Mother is working part time	0.183	-0.010	1720	1713
	(0.387)	(0.014)		
Father is working part time	0.083	0.000	1686	1683
	(0.276)	(0.010)		
Indicator: Dad's education high school or less	0.633	0.020	1743	1738
	(0.482)	(0.018)		
Indicator: Mom's education high school or less	0.528	0.000	1743	1738
	(0.499)	(0.018)		
Father income (thousands)	362.622	2.990	1688	1686
	(226.872)	(8.200)		
Mother income (thousands)	223.054	2.750	1722	1719
	(176.111)	(6.465)		
Father's age at birth of child	33.198	-0.280	1685	1684
	(6.024)	(0.225)		
Mother's age at birth of child	30.220	-0.040	1690	1685
	(4.981)	(0.184)		

Notes: Table shows means for the control families in the TipsByText project. Second column shows results from regressions of characteristics on the randomization indicator. Stars indicate significance levels from associated t-test using cluster-robust standard errors on family level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2

Comparing attrition across control and treatment groups

	Control mean (SD)	Difference between treated and control (SE)	N Treated	N Control
Analysis on full sample				
Attrited	0.143	0.010	1743	1738
	(0.350)	(0.012)		
Moved out of municipality	0.098	0.000	1743	1738
	(0.297)	(0.010)		
Missing language assessment	0.045	0.010	1743	1738
	(0.208)	(0.007)		
Analysis on attrited sample				
Baseline language assessment variables				
Baseline: Total % correct	0.492	0.040**	217	206
	(0.214)	(0.019)		
Baseline: Total % wrong	0.422	-0.040***	217	206
	(0.181)	(0.016)		
Baseline: Total % missing	0.085	0.010	217	206
	(0.159)	(0.014)		
Baseline: Comprehension, % correct	0.646	0.030	217	206
	(0.197)	(0.018)		
Baseline: Vocabulary, % correct	0.419	0.030	217	206
	(0.246)	(0.022)		
Baseline: Rhyme detection, % correct	0.569	0.030	217	206
	(0.343)	(0.028)		
Baseline: Concepts of print, % correct	0.523	0.050*	168	166
	(0.287)	(0.029)		
Baseline: Letter identification, % correct	0.376	0.110**	135	120
	(0.297)	(0.045)		
Baseline: Deletion, % correct	0.449	0.090**	135	120
	(0.315)	(0.040)		
Baseline: Communication, mean	2.844	0.100	208	201
	(0.697)	(0.063)		
Baseline: Age (years) at language test	4.749	0.050*	217	206
	(0.818)	(0.030)		

Panel A: Attrition immediately after program delivery

Notes: Table compares characteristics of parents and children for children without an endline language skills measure across the control and treatment groups Second column shows results from regressions of characteristics on the randomization indicator, preschool fixed effects, and indicators for the test version. Stars indicate significance levels from associated t-test using cluster-robust standard errors on family level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2 continued
Comparing attrition across control and treatment groups

	Control mean (SD)	Difference between treated and control (SE)	N Treated	N Control
Demographic variables				
Indicator: Boy	0.459	0.070	256	244
	(0.499)	(0.049)		
APGAR-score < 10 (max)	0.059	0.010	229	221
	(0.236)	(0.024)		
Low birth weight (< 2500 g)	0.068	-0.030	229	221
	(0.252)	(0.023)		
Number of siblings in intervention	1.324	-0.120	249	241
	(1.212)	(0.116)		
Indicator: Firstborn child (mother)	0.480	0.040	256	244
	(0.501)	(0.049)		
Indicator: One adult in home	0.154	0.050	249	241
	(0.361)	(0.035)		
Indicator: First/second generation immigrant	0.287	-0.030	256	244
	(0.453)	(0.040)		
Mother is working part time	0.176	-0.040	243	238
	(0.382)	(0.036)		
Father is working part time	0.115	0.060	241	234
	(0.320)	(0.034)		
Indicator: Dad's education high school or less	0.611	0.000	256	244
	(0.489)	(0.047)		
Indicator: Mom's education high school or less	0.475	-0.020	256	244
	(0.500)	(0.048)		
Father income (thousands)	361.010	-24.630	241	235
	(235.080)	(23.513)		
Mother income (thousands)	211.534	17.620	245	239
	(194.176)	(20.329)		
Father's age at birth of child	32.869	-0.570	240	234
	(6.163)	(0.653)		
Mother's age at birth of child	30.047	-0.180	240	235
	(5.289)	(0.535)		

Panel A: Attrition immediately after program delivery

Notes: Table compares characteristics of parents and children for children without an endline language skills measure across the control and treatment groups Second column shows results from regressions of characteristics on the randomization indicator, preschool fixed effects, and indicators for the test version. Stars indicate significance levels from associated t-test using cluster-robust standard errors on family level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2 continued
Comparing attrition across control and treatment groups

	Control mean (SD)	Difference between treated and control (SE)	N Treated	N Control
Analysis on full sample				
Attrited	0.271	-0.020	1210	1214
	(0.445)	(0.016)		
Moved out of municipality	0.249	-0.020	1210	1214
	(0.432)	(0.015)		
Missing language assessment	0.022	0.000	1210	1214
	(0.148)	(0.006)		
Analysis on attrited sample				
Baseline language assessment variables				
Baseline: Total % correct	0.462	0.030	271	317
	(0.214)	(0.016)		
Baseline: Total % wrong	0.434	-0.020*	271	317
	(0.186)	(0.013)		
Baseline: Total % missing	0.104	0.000	271	317
	(0.160)	(0.012)		
Baseline: Comprehension, % correct	0.627	0.040**	271	317
	(0.194)	(0.016)		
Baseline: Vocabulary, % correct	0.391	0.040*	271	317
	(0.244)	(0.019)		
Baseline: Rhyme detection, % correct	0.474	0.010	271	317
	(0.361)	(0.026)		
Baseline: Concepts of print, % correct	0.445	0.030	197	224
	(0.273)	(0.025)		
Baseline: Letter identification, % correct	0.397	-0.060	80	80
	(0.314)	(0.064)		
Baseline: Deletion, % correct	0.414	0.070	80	80
	(0.307)	(0.053)		
Baseline: Communication, mean	2.689	0.120**	265	314
	(0.705)	(0.052)		
Baseline: Age (years) at language test	4.337	0.010	271	317
	(0.773)	(0.024)		

Panel B: Attrition at the one-year follow-up

Notes: Table compares characteristics of parents and children for children without an endline language skills measure across the control and treatment groups Second column shows results from regressions of characteristics on the randomization indicator, preschool fixed effects, and indicators for the test version. Stars indicate significance levels from associated t-test using cluster-robust standard errors on family level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2 continued
Comparing attrition across control and treatment groups

	Control mean (SD)	Difference between treated and control (SE)	N Treated	N Control
Demographic variables				
Indicator: Boy	0.484	0.090**	306	349
	(0.500)	(0.042)		
APGAR-score < 10 (max)	0.059	-0.010	283	324
	(0.235)	(0.018)		
Low birth weight (< 2500 g)	0.068	-0.030	283	324
	(0.252)	(0.021)		
Number of siblings in intervention	1.225	-0.030	302	347
	(1.138)	(0.089)		
Indicator: Firstborn child (mother)	0.473	0.000	306	349
	(0.500)	(0.040)		
Indicator: One adult in home	0.153	-0.040	302	347
	(0.360)	(0.029)		
Indicator: First/second generation immigrant	0.249	-0.040	306	349
	(0.433)	(0.033)		
Mother is working part time	0.169	0.000	298	343
	(0.375)	(0.032)		
Father is working part time	0.107	0.000	289	338
	(0.309)	(0.027)		
Indicator: Dad's education high school or less	0.633	-0.060	306	349
	(0.483)	(0.041)		
Indicator: Mom's education high school or less	0.510	-0.030	306	349
	(0.501)	(0.042)		
Father income (thousands)	347.947	19.300	290	338
	(235.127)	(21.331)		
Mother income (thousands)	227.654	11.600	299	346
	(205.230)	(16.360)		
Father's age at birth of child	33.102	-0.040	289	337
	(6.215)	(0.531)		
Mother's age at birth of child	30.254	0.040	291	339
	(5.173)	(0.438)		

Panel B: Attrition at the one-year follow-up

Notes: Table compares characteristics of parents and children for children without an endline language skills measure across the control and treatment groups Second column shows results from regressions of characteristics on the randomization indicator, preschool fixed effects, and indicators for the test version. Stars indicate significance levels from associated t-test using cluster-robust standard errors on family level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3

Comparing attrited group with remaining study population

	Study sample mean (SD)	Difference between attrited and study sample (SE)	N Attrited	N Study sample
Stratification variables				
Indicator: Estimated endline test version = 3	0.035	-0.020**	493	2981
	(0.183)	(0.009)		
Indicator: Estimated endline test version = 4	0.351	0.000	493	2981
	(0.477)	(0.013)		
Indicator: Estimated endline test version = 5	0.335	0.010	493	2981
	(0.472)	(0.015)		
Indicator: Estimated endline test version = School (group)	0.280	0.010	493	2981
	(0.449)	(0.014)		
Baseline language assessment variables				
Baseline: Total % correct	0.475	0.000	423	2786
	(0.201)	(0.010)		
Baseline: Total % wrong	0.427	-0.010	423	2786
	(0.174)	(0.009)		
Baseline: Total % missing	0.099	0.010	423	2786
	(0.155)	(0.008)		
Baseline: Comprehension, % correct	0.655	-0.010	423	2786
	(0.191)	(0.010)		
Baseline: Vocabulary, % correct	0.413	-0.010	423	2786
	(0.238)	(0.012)		
Baseline: Rhyme detection, % correct	0.470	0.000	423	2786
	(0.355)	(0.015)		
Baseline: Concepts of print, % correct	0.473	0.000	335	1759
	(0.265)	(0.015)		
Baseline: Letter identification, % correct	0.361	0.030	255	819
	(0.316)	(0.027)		
Baseline: Deletion, % correct	0.445	-0.010	255	819
	(0.303)	(0.023)		
Baseline: Communication, mean	2.747	-0.050	409	2741
	(0.648)	(0.035)		
Baseline: Age (years) at language test	4.304	0.010	423	2786
	(0.876)	(0.015)		

Panel A: Attrition immediately after program delivery

Notes: Table compares characteristics of parents and children for children without an endline language skills measure with the study population. Second column shows results from regressions of characteristics on the randomization indicator, preschool fixed effects, and indicators for the test version. Stars indicate significance levels from associated t-test using cluster-robust standard errors on family level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3 continued
Comparing attrited group with remaining study population

	Study sample mean (SD)	Difference between attrited and study sample (SE)	N Attrited	N Study sample
Demographic variables				
Indicator: Boy	0.511	-0.020	500	2981
	(0.500)	(0.026)		
APGAR-score < 10 (max)	0.065	-0.010	450	2858
	(0.246)	(0.013)		
Low birth weight (< 2500 g)	0.055	0.010	450	2858
	(0.229)	(0.014)		
Number of siblings in intervention	1.289	-0.050	490	2963
	(1.067)	(0.061)		
Indicator: Firstborn child (mother)	0.431	0.060**	500	2981
	(0.495)	(0.026)		
Indicator: One adult in home	0.114	0.040**	490	2963
	(0.318)	(0.019)		
Indicator: First/second generation immigrant	0.198	0.040**	500	2981
	(0.398)	(0.021)		
Mother is working part time	0.180	-0.020	481	2952
	(0.384)	(0.020)		
Father is working part time	0.078	0.050***	475	2894
	(0.269)	(0.017)		
Indicator: Dad's education high school or less	0.648	-0.040	500	2981
	(0.478)	(0.025)		
Indicator: Mom's education high school or less	0.538	-0.060**	500	2981
	(0.499)	(0.026)		
Father income (thousands)	365.704	-18.090	476	2898
	(222.294)	(12.046)		
Mother income (thousands)	225.058	-9.030	484	2957
	(171.373)	(9.881)		
Father's age at birth of child	33.095	-0.390	474	2895
	(6.029)	(0.347)		
Mother's age at birth of child	30.234	-0.400	475	2900
	(4.915)	(0.282)		

Panel A: Attrition immediately after program delivery

Notes: Table compares characteristics of parents and children for children without an endline language skills measure with the study population. Second column shows results from regressions of characteristics on the randomization indicator, preschool fixed effects, and indicators for the test version. Stars indicate significance levels from associated t-test using cluster-robust standard errors on family level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3 continued
Comparing attrited group with remaining study population

	Study sample mean (SD)	Difference between attrited and study sample (SE)	N Attrited	N Study sample
Stratification variables				
Indicator: Estimated endline test version = 3	0.050	0.000	651	1769
	(0.219)	(0.012)		
Indicator: Estimated endline test version = 4	0.549	-0.030	651	1769
	(0.498)	(0.018)		
Indicator: Estimated endline test version = 5	0.386	0.010	651	1769
	(0.487)	(0.016)		
Indicator: Estimated endline test version = School (group)	0.014	0.030***	651	1769
	(0.118)	(0.008)		
Baseline language assessment variables				
Baseline: Total % correct	0.434	0.020*	588	1629
	(0.193)	(0.010)		
Baseline: Total % wrong	0.438	-0.030***	588	1629
	(0.171)	(0.009)		
Baseline: Total % missing	0.128	0.000	588	1629
	(0.172)	(0.009)		
Baseline: Comprehension, % correct	0.633	0.010	588	1629
	(0.196)	(0.010)		
Baseline: Vocabulary, % correct	0.373	0.020**	588	1629
	(0.233)	(0.012)		
Baseline: Rhyme detection, % correct	0.350	0.030	588	1629
	(0.320)	(0.016)		
Baseline: Concepts of print, % correct	0.369	0.050***	421	694
	(0.233)	(0.018)		
Baseline: Letter identification, % correct	0.243	0.160*	160	25
	(0.267)	(0.087)		
Baseline: Deletion, % correct	0.336	0.070	160	25
	(0.273)	(0.066)		
Baseline: Communication, mean	2.602	0.010	574	1617
	(0.628)	(0.034)		
Baseline: Age (years) at language test	3.804	0.030**	583	1628
	(0.615)	(0.016)		

Panel B: Attrition at the one-year follow-up

Notes: Table compares characteristics of parents and children for children without an endline language skills measure with the study population. Second column shows results from regressions of characteristics on the randomization indicator, preschool fixed effects, and indicators for the test version. Stars indicate significance levels from associated t-test using cluster-robust standard errors on family level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3 continued
Comparing attrited group with remaining study population

	Study sample mean (SD)	Difference between attrited and study sample (SE)	N Attrited	N Study sample
Demographic variables				
Indicator: Boy	0.508	-0.010	655	1769
	(0.500)	(0.027)		
APGAR-score < 10 (max)	0.068	-0.020*	607	1708
	(0.252)	(0.013)		
Low birth weight (< 2500 g)	0.056	-0.010	607	1708
	(0.229)	(0.014)		
Number of siblings in intervention	1.224	-0.090	649	1756
	(1.059)	(0.058)		
Indicator: Firstborn child (mother)	0.428	0.070***	655	1769
	(0.495)	(0.026)		
Indicator: One adult in home	0.109	0.030	649	1756
	(0.312)	(0.019)		
Indicator: First/second generation immigrant	0.201	0.000	655	1769
	(0.401)	(0.021)		
Mother is working part time	0.178	0.000	641	1752
	(0.383)	(0.020)		
Father is working part time	0.075	0.040**	627	1717
	(0.263)	(0.017)		
Indicator: Dad's education high school or less	0.649	-0.050*	655	1769
	(0.477)	(0.026)		
Indicator: Mom's education high school or less	0.531	-0.020	655	1769
	(0.499)	(0.026)		
Father income (thousands)	362.686	-11.730	628	1720
	(211.125)	(12.818)		
Mother income (thousands)	219.601	11.670	645	1753
	(167.529)	(9.740)		
Father's age at birth of child	33.025	-0.100	626	1720
	(6.114)	(0.346)		
Mother's age at birth of child	30.126	-0.060	630	1720
	(4.889)	(0.279)		

Panel B: Attrition at the one-year follow-up

Notes: Table compares characteristics of parents and children for children without an endline language skills measure with the study population. Second column shows results from regressions of characteristics on the randomization indicator, preschool fixed effects, and indicators for the test version. Stars indicate significance levels from associated t-test using cluster-robust standard errors on family level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4

Effects of TipsByText on probability of falling into various quartiles of the language assessment at endline

	Dependent variable: Probability of falling in different quartiles of endline language distribution			
	(1) 1st quartile	(2) 2nd quartile	(3) 3rd quartile	(4) 4th quartile
Treatment	0.007	0.015	-0.017	-0.005
	(0.013)	(0.016)	(0.016)	(0.014)
Preschool FE	✓	✓	✓	✓
Test version FE	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓
SE cluster level	Family	Family	Family	Family
Observations	2977	2977	2977	2977
R^2	0.392	0.070	0.083	0.304

Panel A: Outcomes measured immediately after program delivery

	Dependent variable: Probability of falling in different quartiles of endline language distribution			
	(1) 1st quartile	(2) 2nd quartile	(3) 3rd quartile	(4) 4th quartile
Treatment	-0.019	0.005	0.043**	-0.034*
	(0.018)	(0.020)	(0.021)	(0.018)
Preschool FE	✓	✓	✓	✓
Test version FE	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓
SE cluster level	Family	Family	Family	Family
Observations	1763	1763	1763	1763
R^2	0.298	0.102	0.106	0.292

Panel B: Outcomes measured at the one-year follow-up

Notes: Table presents regressions of indicators for falling into quartiles of the standardized outcome on the treatment indicator, preschool fixed effects and test version, the standardized baseline measure, and an indicator for whether baseline was missing. Singletons in fixed effects are dropped. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5

Effects of TipsByText on alternative versions of the outcome

	(1) Standardized index	(2) Proportion correct	(3) Proportion wrong	(4) Proportion missing	(5) Δ Prop. correct
Treatment	-0.029	-0.004	0.004	0.001	-0.001
	(0.026)	(0.005)	(0.005)	(0.003)	(0.005)
Mean of treated group	-0.048	0.570	0.377	0.053	0.108
S.D. of treated group	1.040	0.208	0.179	0.112	0.121
N treated group	1484	1484	1484	1484	1381
Mean of control group	-0.000	0.578	0.370	0.051	0.109
S.D. of control group	0.999	0.207	0.177	0.111	0.150
N control group	1493	1493	1493	1493	1400
Preschool FE	✓	✓	✓	✓	✓
Test version FE	✓	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓	
SE cluster level	Family	Family	Family	Family	Family
Observations	2977	2977	2977	2977	2781
R^2	0.554	0.615	0.504	0.329	0.205

Panel A: Outcomes measured immediately after program delivery

	(1) Standardized index	(2) Proportion correct	(3) Proportion wrong	(4) Proportion missing	(5) Δ Prop. correct
Treatment	0.000	0.001	-0.002	-0.001	0.001
	(0.035)	(0.006)	(0.006)	(0.003)	(0.008)
Mean of treated group	-0.006	0.613	0.345	0.042	0.194
S.D. of treated group	0.988	0.191	0.167	0.095	0.164
N treated group	902	902	902	902	803
Mean of control group	-0.000	0.619	0.341	0.039	0.194
S.D. of control group	1.000	0.194	0.174	0.089	0.173
N control group	861	861	861	861	783
Preschool FE	✓	✓	✓	✓	✓
Test version FE	✓	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓	
SE cluster level	Family	Family	Family	Family	Family
Observations	1763	1763	1763	1763	1586
R^2	0.486	0.563	0.467	0.413	0.259

Panel B: Outcomes measured at the one-year follow-up

Notes: Table presents regressions of versions of the outcome Table presents regressions of the standardized outcome on the treatment indicator, preschool fixed effects and test version, the standardized baseline measure, and an indicator for whether baseline was missing. Singletons in fixed effects are dropped. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Economics Working Papers

2021-07	Johannes Schünemann, Holger Strulik and Timo Trimborn: Optimal Demand for Medical and Long-Term Care
2021-08	Tom Lane, Daniele Nosenzo and Silvia Sonderegger: Law and Norms: Empirical Evidence
2021-09	Ina C. Jäkel: Export Credit Guarantees: Direct Effects on the Treated and Spillovers to their Suppliers
2021-10	Martin Paldam: Measuring Democracy - Eight indices: Polity, Freedom House and V-Dem
2021-11	Nicola Maaser and Thomas Stratmann: Costly Voting in Weighted Committees: The case of moral costs
2021-12	Martin M. Andreasen, Giovanni Caggiano, Efrem Castelnuovo and Giovanni Pellegrino: Why Does Risk Matter More in Recessions than in Expansions?
2021-13	Søren Albeck Nielsen: How to Cope with Dyslexia: The Effects of Special Education on Academic Performance, Personality Traits, and Well-being
2022-01	Michael Koch and Ilya Manuylov: Measuring the Technological Bias of Robot Adoption and its Implications for the Aggregate Labor Share
2022-02	Olga Balakina, Claes Bäckman, Andreas Hackethal, Tobin Hanspal and Dominique M. Lammer: Good Peers, Good Apples? Peer Effects in Portfolio Quality
2022-03	Kazuhiko Sumiya and Jesper Bagger: Income Taxes, Gross Hourly Wages, and the Anatomy of Behavioral Responses: Evidence from a Danish Tax Reform
2022-04	Hans Schytte Sigaard: Labor Supply Responsiveness to Tax Reforms
2022-05	Edi Karni and Marie-Louise Vierø: Comparative Incompleteness: Measurement, Behavioral Manifestations and Elicitation
2022-06	Marie-Louise Vierø: Lost in objective translation: Awareness of unawareness when unknowns are not simply unknowns
2022-07	Daniele Nosenzo, Erte Xiao and Nina Xue: Norm-Signalling Punishment
2023-01	Susanna Loeb, Michala Iben Riis-Vestergaard and Marianne Simonsen: Supporting Language Development through a Texting Program: Initial Results from Denmark