From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Psychology

2021

TipsByText: A Parent-Oriented Language Intervention for Preschool Children

Riis-Vestergaard, Michala Iben

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01pv63g3376

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

TipsByText: A Parent-Oriented Language Intervention for Preschool Children

Michala Iben Riis-Vestergaard

a dissertation
presented to the faculty
of princeton university
in candidacy for the degree
of Doctor of Philosophy

recommended for acceptance
by the Department of
Psychology

Adviser: Johannes Haushofer

September 2021

© Copyright by Michala Iben Riis-Vestergaard, 2021. All rights reserved.

Abstract

Early-life skills are a strong predictor of later life outcomes, and parents play an instrumental role in their children's acquisition of early-life skills. However, there is still a question in the literature about the effectiveness of leveraging parents to improve children's early-life skills as a large-scale policy tool. In this dissertation, we estimate the impact of a parent-oriented intervention, TipsByText, in a naturalistic setting with a large and diverse sample. TipsByText aims to increase language skills among preschool children by sending text messages directly to their parents with tips about how to incorporate simple language stimulation exercises into their everyday lives. To evaluate the effectiveness of the TipsByText intervention, we use data from a randomized controlled trial (RCT) conducted in Denmark in 2019/2020 among 3,305 families with 3,804 three-to-five-year-old children. We estimate that the TipsByText intervention has a tightly estimated null effect on children's language skills in the short run, indicating that the TipsByText intervention was not successful in improving language skills in our sample. We draw on various qualitative and quantitative data sources to analyze potential explanations for why the TipsByText intervention did not improve children's language skills. As none of the available data sources provide us with a conclusive explanation for the lack of positive effect of the TipsByText intervention, a potential implication of the current study is that leveraging parents to improve children's early-life skills might not be as effective as a large-scale policy tool as the literature suggests.

Contents

Abstract **iii**

0 Introduction **1**

1 The TipsByText intervention and its effect on children's language skills **11**
- 1.1 The TipsByText intervention . . . 12
- 1.2 Experimental details . . . 18
- 1.3 Results . . . 34
- 1.4 Conclusion . . . 48

2 Why did TipsByText not improve children's language skills? **51**
- 2.1 Exploring the link between the intervention and parent behavior . . . 52
- 2.2 Exploring the link between parent behavior and children's language skills . . . 69
- 2.3 Conclusion . . . 78

3 Discussion **82**
- 3.1 Limitations of the current study . . . 84
- 3.2 Possible positive impact beyond improving children's language skills . . . 90
- 3.3 Future directions . . . 94

References **100**

Appendix A Additional Analyses **101**

Appendix B Language Assessment Details **124**

The following authors contributed to the content of this dissertation: Michala Iben Riis-Vestergaard (first author) conducted all analyses and wrote the first draft, Marianne Simonsen (PI) supervised and provided detailed comments on analyses and writing, and Helene Bie Lilleør and Susanna Loeb reviewed and approved the analyses.

Listing of figures

1.1 Text messages from an example week 15
1.2 Map of participating municipalities 21
1.3 Sample overview 26
1.4 RCT timeline 29
1.5 Graphical representation of the treatment effect on children's language skills 38
1.6 Distribution of language skills in treatment vs. control group 42
1.7 Overview of results in Chapter 1 50

2.1 Theory of change for the TipsByText intervention 52
2.2 Text message delivery 55
2.3 Intervention enrollment 56
2.4 Dose response 58
2.5 Parent survey responses 60
2.6 Differential language growth for the engaged 72

A.1 Power simulations 102

Listing of tables

1.1 TipsByText language domains: Overview 19
1.2 TipsByText language domains: Examples 20
1.3 Overview of sample across collaborating municipalities 22
1.4 Overview of subtests in the language assessment 28
1.5 Baseline balance: Language skills 31
1.6 Baseline balance: Socioeconomic variables 32
1.7 Attrition 35
1.8 Treatment effect on the standardized outcome 37
1.9 Treatment effects on components on index 40
1.10 Differential treatment effects 45

2.1 Parent survey responses to the free text question 61
2.2 Parent engagement rubric 64

A.1 Representativeness of sample within collaborating municipalities 103
A.2 Baseline balance: Stratification variables 104
A.3 Differential attrition among those children who moved 105
A.4 Differential attrition among those children who did not move 106
A.5 Baseline balance: Language skills 107
A.6 Baseline balance: Socioeconomic variables 108
A.7 Representativeness of the evaluation sample out of the full sample 109
A.8 Treatment effect on different definitions of the outcome measure 110
A.9 Probability for being in different quartiles of outcome distribution 111
A.10 Treatment effect for different subgroups 112
A.11 Treatment effect on the treated 113
A.12 Characteristics of children whose parents opted out 114
A.13 Additional categories of parent's free text survey responses 115
A.14 Characteristics of the engaged vs. children of parents who opted out 116
A.15 Characteristics of the engaged vs. children of parents who did not opt out . 117

A.16 Characteristics of the engaged vs. the control group 118
A.17 Language growth for the engaged vs. children of parents who opted out . . 119
A.18 Language growth for the engaged vs. children of parents who did not opt out 120
A.19 Language growth for the engaged vs. children in the control group 121
A.20 Treatment effect by child age 122
A.21 Treatment effect by baseline language skills 122
A.22 Treatment effect by parental labor market participation 123

To Christina Hall Frederiksen;
My high school math teacher who taught me the beauty of math
and the value of truly heartfelt mentorship

Acknowledgments

Thank you to Johannes Haushofer for believing in me from our first meeting (when you nodded enthusiastically at me when I asked a question at a seminar), for giving me the fantastic opportunity of doing a PhD at Princeton University, and for helping me land my dream job at J-PAL. To Casey Lew-Williams for all your help and feedback and for being the best DGS a graduate student could ever wish for. To Diana Tamir, Elke Weber, and Adele Goldberg for providing valuable feedback on this dissertation. To the Joint Degree Program team, including Susan Fiske and Sara McLanahan, for expanding my horizons. To David Dreyer Lassen for opening a door for me, and for encouraging me to go through another door when the opportunity arose. To Marianne Simonsen for asking and answering critical questions and for providing detailed feedback on this dissertation. To Helene Bie Lilleør for giving me the experience I didn't know I needed to get unstuck, and for trusting me with the responsibility of conducting a high-quality evaluation of TipsByText. To Signe Haarh Petersen and Mikkel Stahlschmidt for invaluable research assistance, and to Johanne Bergmann and Emilie Dencker for keeping me on my toes. To all my wonderful colleagues at the ROCKWOOL Foundation who educated me in the complexity of truly interdisciplinary work and who taught me to write ugly notes on post-its. To my family, Marianne, Søren, and Simone, for loving me unconditionally, for steadying me every time I tripped, and for catching me the one time I did actually fall. To Dorte and Erik, for your unwavering belief in me and what I am capable of achieving. To Katja, for being my best friend and for philosophizing with me about everything in life from the value of a PhD to catching whales with cranes. To my life coach, Milton, for keeping me sane, for playing with me, and for encouraging mindfulness at all times. And to Emil for... everything. Including loving me enough to let me go to Denmark for two years. And for enduring my endless stream of questions about grammar, LaTeX, and Stata while being forced to share an office with me during the lockdown. Without your encouragement and support, this dissertation would not have existed. You are the best thing that ever happened to me, and I cannot wait to embark on this new life adventure with you.

0

Introduction

Early-life skills are a strong predictor of later life outcomes, such as academic achievement and earnings (e.g. Cunha et al., 2006). In fact, the societal return on investments, in terms of increased earnings and decreased costs associated with, e.g., social transfers and incarcerations, is much higher for young children than for older children or adults (Heck-

man, 2008). For this reason, there is a great policy focus on improving early life skills as a goal in itself and as a way to close skills gaps related to, e.g., socioeconomic status (e.g. National Early Literacy Panel's report, 2009).

Parents play an instrumental role in their children's acquisition of early-life skills, including early language development. A large literature in developmental psychology and sociology has identified strong correlations between how and how much parents talk to their children and children's early language development (e.g., Schwab and Lew-Williams, 2016; Gilkerson et al., 2018; Huttenlocher et al., 1991). Early research in this field, starting with the seminal work of Betty Hart and Todd Risley (1995), colloquially called the "30 Million Word Gap" study, focused largely on the socioeconomic gradient in child-directed speech (e.g., Rowe, 2008; Farkas and Beron, 2004; Hoff-Ginsberg, 1991; Fernald et al., 2013). However, research has found that that parents' behavior predicts children's early language development beyond socioeconomic variables (e.g., Brooks-Gunn and Markman, 2005; Hirsh-Pasek et al., 2015; Hoff, 2003; Rowe, 2018). Involving parents as intervention agents, therefore, seems to be a promising vehicle in the policy goal of improving early life skills and decrease skills gaps (Hoff, 2013; Leffel and Suskind, 2013).

Yet, despite the promise of parent-oriented interventions for improving early skills development, there is limited causal evidence on the effectiveness of scalable parent-oriented interventions. There are two categories of parent-oriented interventions that have been studied rigorously and 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 limitations.

Shared-reading interventions, in which parents are encouraged to read with their chil-

dren and to interact with the children around the shared reading material, have been widely implemented in the US with programs such as "Reach Out and Read" and "Reading is Fundamental" (e.g. Canfield et al., 2020). In an extensive meta-analysis on existing language-enhancing programs that have been evaluated using experimental or quasi-experimental methods, National Early Literacy Panel (2009) identified 19 randomized or quasi-randomized studies of the impacts of shared-reading interventions. Children in these studies were exposed to an increase in shared reading by either parents or teachers, and/or to a change in the style of shared reading activities. The report concludes that shared-reading interventions have positive and moderate effects on children's oral language and print knowledge. Thus, shared reading interventions have been widely implemented and can be successful in increasing children's early language skills. This class of interventions has the benefits that they are cheap to implement and easy to scale (e.g. through healthcare providers, as is the case with the "Reach out and Read" program). However, a weakness of this type of intervention is that it focuses 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.

The other type of intervention that has been studied rigorously and been found to have consistent positive effects on children's early language skills is parent education programs, in which parents receive training in effective language stimulation strategies either through a home visiting program or a group training program in a classroom setting. One widely studied example of such a program is the program "It Takes Two to Talk" from The Hanen Centre, in which parents of children with identified language delays are taught practical strategies to help their children learn language by a speech-language pathologist through a

series of small-group sessions home visits (Pepper et al., 2004). In a meta-analysis of group-based parent-implemented language interventions for children with any type of language impairment, Roberts and Kaiser (2011) identified 18 studies (of which eight focused on the Hanen Parent Program) that included at least one language outcome measure. Parents in these interventions were taught strategies that included responding to child communication, balancing adult–child turns, and using language models; the interventions lasted between 10 and 18 weeks, and 11 of the interventions included home visits as well as group training. The authors found that in general, the group-based parent-implemented language interventions had positive effects on children's language outcomes, with positive and significant meta-effect sizes across six of the seven language outcome constructs included across the 18 studies. The benefit of this type of program is that it typically covers many aspects of children's language development, in contrast to the shared reading interventions. However, its weakness is that it is costly to implement and difficult to scale: Home-visiting programs are time intensive on the part of the visitor (e.g., a nurse or trained language specialist) and are, therefore, costly to implement from a policy perspective. Classroom-type parent education programs are costly for parents to attend, in that they need to show up at a given place and time. This can result in high attrition rates, particularly among lower resource families, who might be expected to benefit the most from the programs.

Thus, there is evidence from the literature that parent-oriented interventions can increase children's early language skills. However, the most studied types of parent-oriented interventions seem to face a trade-off between scope and scalability: Either they are relatively cheap and easy to scale but narrow in scope, such as the shared-reading intervention, or they are extensive in scope but expensive to implement and difficult to scale, such as the

parent-education programs.

Furthermore, most of the studies on the effectiveness of parent-implemented interventions are efficacy studies in which the intervention's impact is tested in a small scale study under ideal settings (e.g., including heavy involvement of the researchers, Bleses et al., 2018b). For example, in a recent meta-analysis focusing on experimental evidence on the effectiveness of parent-oriented language interventions (defined broadly as interventions in which "parents received explicit training to implement specific language-facilitating strategies with their children") for children at-risk for language impairments due to low SES, Heidlage et al. (2020) identified 25 randomized controlled trials examining linguistic outcomes for young children. The authors found significant positive meta-effects of parent-oriented interventions on children's language skills, particularly on expressive vocabulary and expressive language. However, most studies included in the analysis are based on very small evaluation sample sizes, ranging from $N = 15$ to $N = 225$ with a median of $N = 41$, across the treatment and control groups. The authors thus conclude that "There is a continuing need for high quality studies examining the effects of parent-implemented language intervention." Similarly, the 18 studies included in Roberts and Kaiser (2011) were based on sample sizes varying from $N = 12$ to $N = 152$ with a mean of $N = 38$. And in their meta-analysis of 32 parent-oriented language-enhancing programs, National Early Literacy Panel (2009) conclude that "there were significant problems with the quality of much of the research in this area," as many studies used simple pretest-posttest designs instead of experimental methods.

In this project, we try to solve the problems of (1) a trade-off between scope and scalability in parent-oriented interventions, and (2) a lack of knowledge of the effectiveness of

parent-oriented interventions in large-scale naturalistic settings. We do this by testing the impact of a parent-oriented intervention, TipsByText, 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.

TipsByText aims to increase language skills among preschool children by sending text messages directly to their parents with tips about how to incorporate simple language stimulation exercises into their everyday lives. Thus, the intervention addresses the trade-off between scope and scalability in the following ways. First, the intervention covers a wide range of tips for how to engage children in active language stimulation, including but not limited to shared reading. Second, the intervention is flexible and imposes very little time cost on the parent, as the text messages are short and all tips include language stimulation exercises that can be implemented in the busy everyday lives of families with small children. Third, the intervention is inexpensive for implementers and highly scalable, as the marginal cost of sending text messages to additional parents is negligible. The content of the text messages is grounded in psychology and behavioral science theory and targets some of parents' behavioral barriers to active language stimulation: lack of information, cognitive load, and lack of motivation.

To evaluate the effectiveness of the TipsByText intervention, we use data from a randomized controlled trial (RCT) conducted in Denmark in 2019/2020. Parents in the treatment group received the intervention for eight months, while parents in the control group received no intervention (treatment as usual). The RCT is based on a sample of 3,305 families with 3,804 three-to-five-year-old children from five (out of 98 total) different and geographically dispersed municipalities in Denmark. Children's language skills are assessed

using the Danish language assessment "Sprogvurdering 3-6" both before and after the intervention.

The experiment and data set, to which we have primary access, have several advantages relative to previous impact evaluations of parental programs. First, the experimental design improves the external validity of the estimated effects compared to previous studies, which often have small and convenient samples and opt-in designs. Specifically, the present experiment is based on a large sample, which allows us to estimate even small effects (down to 0.1 standard deviations [S.D.]) with tight standard errors. The experiment is based on an opt-out design, which is often increases the take-up of an intervention compared to an opt-in design, because it changes the default (e.g., Davidai et al., 2012). Indeed, by the end of the intervention, 86% of the children in the treatment group still had at least one parent enrolled in the program. In turn, high take-up leads to a high diversity of the treated group, which improves the external validity of the estimated effects of the intervention along different dimensions, such as baseline language skills and parental education and income.

Second, we can link all children and parents in the sample to an extensive database of administrative records on their demographics, education, and income. The link to administrative data on family outcomes improves the accuracy of the estimate of the effect size compared to previous studies that rely on self-reported data. Specifically, the administrative data are extensive, almost complete, and have very low or no measurement error, which decreases standard errors and the likelihood that the estimated effect is biased toward zero. The administrative data also allows us to follow the children who participate in the RCT into the future to estimate the long-term effects of the intervention.

This dissertation has two aims presented in two separate chapters:

Aim 1: Determine whether the parent-oriented intervention, TipsByText, is successful in improving the language skills of preschool children and estimate the effect size of the program.

Aim 2: Understand what drives the effect of the intervention or, in case there is no effect, understand what hinders the effect.

In Chapter 1, we show that the TipsByText program did not succeed in improving children's early language skills as measured by the Danish language assessment, "Sprogvurdering 3-6." The estimated effect size is extremely close to zero across multiple different empirical specifications, and the null effect is very tightly estimated. We find no evidence of a positive effect on any of the dimensions of language skills measured with "Sprogvurdering 3-6," and we find no evidence of positive effects for any well-powered subgroups. We thus conclude with a high degree of confidence that the TipsByText intervention did not improve children's average language skills as measured by the language assessment, "Sprogvurdering 3-6."

In Chapter 2, we discuss possible reasons for why there is no positive effect of the TipsByText intervention in our sample. We rule out that the lack of effect is driven by (a) a lack of power, (b) that too few parents received the text messages, (c) or that too many parents opted out of the intervention. We present data from a parent survey and a qualitative evaluation which both indicate that parents appreciated and used the program, and that some parents used the tips in a way that the literature has identified to be beneficial for language stimulation: By dedicating more time to reading with their children, creating discussions about the reading material, increasing the complexity of the vocabulary they used with their children, and conversing more and in more diverse ways with their children. Thus, lack

of compliance or parental engagement with the intervention is unlikely to fully account for the lack of positive effect of the intervention on children's language skills. This conclusion is strengthened by exploratory analyses suggesting that the children of parents who indicated engagement with the intervention did not experience a higher language growth between the baseline and endline language assessment than did other children.

As none of the available data sources provide us with a conclusive explanation for the lack of positive effect of the TipsByText intervention, we are left with two possible broader and potentially overlapping explanations. The first is that the lack of positive effect of the intervention is not caused by a single factor but by a multitude of different biases toward zero, such as many parents not engaging with the intervention, parents not changing their behavior enough to impact their children's language development, the children being too old to experience large benefits of the way their parents speak to them, etc. The second explanation, which can follow directly from the first, is that children's language skills are not easily changed with a light-touch parent-oriented intervention when it is implemented in a large-scale effectiveness study with a large and diverse sample. This interpretation suggests that leveraging parents to improve children's early life skills might not be as effective as a large-scale policy tool as the literature suggests.

The current study was designed to ask whether a particular parent-oriented intervention, TipsByText, could be an effective large-scale policy tool for improving children's early language skills. The evaluation showed that the intervention did not, in fact, improve children's early language skills despite being implemented and used as expected. Thus, this study does not provide evidence on the general effectiveness of parent-oriented interventions. However, it does provide a valuable example of a well-designed and thoroughly

evaluated parent-oriented intervention that fails to create the hypothesized effects. More research is needed before we can conclusively understand whether leveraging parents as a large-scale policy tool can be effective for improving children's early language skills.

"When I see the message, I get so happy. Because when I pick up my daughter, she wants something new every day. But when I don't have anything new, she says: "mama, it's too boring today! I Want Something New.""

– Inaya, recipient of the TipsByText intervention.

1

The TipsByText intervention and its effect on children's language skills

The aim of this chapter is to determine whether the parent-oriented intervention, TipsByText, is successful in improving the language skills of kindergarten children and to

estimate the effect size of the intervention. In this chapter, we will test two main hypotheses:

> *Hypothesis 1.1 (main effect)*: The TipsByText intervention improves the average language skills of children whose parents receive the treatment.
>
> *Hypothesis 1.2 (differential treatment effect)*: The TipsByText intervention decreases early skills gaps by having larger effects among children who score lower in the baseline language assessment and children from lower socioeconomic backgrounds.

All analyses in this chapter will be based on data from a large-scale randomized controlled trial (RCT) conducted in Denmark in 2020 covering 3,305 families with 3,804 three-to-five-years-old children.

1.1 The TipsByText intervention

TipsByText is a text-messaging program that aims at helping parents support their preschool children's language development. The program is based on the philosophy that parents want what is best for their children, and that they want to contribute to their children's learning and development, but that everyday life as a parent with young children is complex and filled with chores. Thus, a fundamental aim of TipsByText is to make it easier for parents to support their children's language development (ROCKWOOL Foundation, 2019).

For eight months, parents receive three weekly text messages with tips for fun and simple language-stimulating games and activities they can do with their children. The text messages are written in a 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 kindergarten/preschool. The curriculum is aimed at parents of three- to five-year-old children, as children in Denmark typically enter kindergarten around the age of three and stay there until they start preschool ("kindergarten class") in the year they turn six.

TipsByText is adapted to the Danish context from an 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 (see https://ready4k.parentpowered.com for more information), and which has shown promising effects on parent behavior and children's language skills in a US context (York et al., 2019).

1.1.1 Design principles behind TipsByText

TipsByText is designed around three core principles aimed at increasing parents' motivation and engagement with the intervention:[1]

- *Fun and easy*: The activities have to be fun, easy to do, and easy to integrate into the daily routines of families with young children. It is important that the text messages are short, affirmative, and reduce the complexity of information. The activities must be motivating and readily accessible to ensure that the parents both will and can use them.

- *No lecturing*: It is at the parents' discretion how they use the text messages and no

[1] These principles are translated almost verbatim from a short report on the adaptation process of the TipsByText program published by the ROCKWOOL Foundation (ROCKWOOL Foundation, 2021)

one monitors how or how much they do so.

- *Practice makes perfect*: The text messages should be based on and promote a growth mindset (Dweck, 2008). An important message is that the parents should practice and do something instead of holding back in fear of not doing the right thing or not doing it right. The messages should be affirmative and give the parents a sense of being acknowledged for their effort.

Furthermore, TipsByText is designed to overcome three behavioral barriers to involved parenting. First, some parents may have insufficient information about the impact their child-directed speech has on their child's language development. To overcome this barrier, each Monday parents receive a "FACT" text designed to instill new knowledge by highlighting the importance of particular skills and parents' role in developing that skill. Second, 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 overcome this barrier, each Wednesday the parents receive a "TIP" text, 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, in order to motivate parents and keep them engaged in the language support of their children, each Friday the parents receive a "GROWTH" text, which provides encouragement and a deeper insight into the benefits of the exercise given on Wednesday (see Figure 1.1 for examples of the FACT, TIP, and GROWTH messages from a given week).

Figure 1.1: Text messages from an example week

	Monday: FACT	**Wednesday: TIP**	**Friday: GROWTH**
Explanation	On Mondays, parents receive a "FACT" text designed to generate knowledge by highlighting the importance of particular skills and the parent's role in developing that skill.	On Wednesdays, parents receive a "TIP" text, which contains ideas for specific activities to incorporate a language-stimulating element into the daily routine. The tip builds on Monday's fact text.	On Fridays, parents receive a "GROWTH" text, which provides encouragement and a deeper insight into the benefits of the exercise given on Wednesday. It often also includes an additional activity.
Example	Children need to know that letters make up words. Research shows that kids with good letter knowledge become good readers.	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?	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?

1.1.2 Adaptation of TipsByText to the Danish context

TipsByText was adapted to the Danish context by The ROCKWOOL Foundation in collaboration with the original designer of the program, Susanna Loeb, and with Danish language researcher and consultant Pia Thomsen, as well as language consultants, parents, and kindergarten leaders and pedagogues from selected municipalities.

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 translation into Danish): an adaptation to the Danish learning culture, and an expansion of the curriculum to include more domains of early language development to make it cover more of the content that 3- to 5-year-old children are expected to master. The core design principles were preserved.

Danish learning culture

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 child care for 3–5-year-old children (or "preschool") is to prepare children for primary school. Thus, cognitive development and the acquisition of knowledge and skills are 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)

On the contrary, the Danish view of learning 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 tradition, the role of child care of 3–5-year-olds, or "børnehaver" (translated to "gardens for children" or "kindergartens" in German), 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 kindergartens, as well as in Danish families, learning happens primarily through play, interaction, activity, and personal investigation. As examples of this view, the "adults" in the Danish kindergartens are not called "teachers" but "pedagogues," and the kindergartens are structured as, and are considered to be, the children's second homes (Kragh-Müller, 2017; OECD, 2006). In fact, interviews with pedagogues in Denmark have revealed resistance to using terms such as "schools, classrooms, teachers, and learning," whereas these terms are included in the common discourse in the US (Kragh-Müller, 2017; Kragh-Müller and Isbell, 2011). To keep salient the differences

in learning traditions between the US and Denmark, we will subsequently refer to Danish child care for 3–5-year-olds as "kindergartens" and to pedagogically trained adults within these kindergartens as "pedagogues."

As research conducted in Danish kindergartens 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 messages 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 (Heidlage et al., 2020; Golinkoff et al., 2006; **?**).

The following adaptation of a GROWTH message provides a good example of the adaption from a focus on formal to more informal learning while keeping the content and learning goal constant. It also shows the emphasis on a growth mindset:

> *US version*: "By teaching your child which direction we read, you're preparing him/her 4K. Ask your child: do we read from left to right, or from right to left?" (York et al., 2019, online appendix)

> *Danish version*: "Showing your child the reading direction helps them learn to read later on. Ask: Can you show me where to start reading? It's ok if your child doesn't know yet. Try again tomorrow. What matters is that you have fun when you read!"

Curriculum

Whereas the US curriculum focuses primarily on supporting early literacy, the Danish curriculum has been expanded to focus on three components of language development, which 3- to 5-year-old children in Denmark are expected to master, and which have been identified in the literature as being important components of children's early language development:

- Vocabulary (formal language: understanding and producing words)
- Early literacy (letters, reading direction, etc.)
- Language use/communication (conversation, dialogue, narrativity, taking turns and initiatives)

Table 1.1 shows the curriculum's focus within each of these domains, and Table 1.2 shows the text messages from an example week for each of these domains.

More information about the adaptation process is available in a short report published by the ROCKWOOL Foundation (ROCKWOOL Foundation, 2021).

1.2 Experimental details

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.

Table 1.1: Language domains addressed by the Danish TipsByText curriculum

Language domain	Language subdomains
Early literacy	1. Rhyme detection 2. Phonological awareness 3. Print-concept knowledge 4. Reading direction 5. Experiences with print 6. Letter identification 7. Deletion 8. Auditory discrimination
Vocabulary	1. Language comprehension (understanding a message) 2. Productive vocabulary (nouns, verbs, adjectives, prepositions)
Communication	1. Taking initiatives 2. Taking turns (you-me-sequences) 3. Maintaining topic in brief conversations 4. Speaking coherently 5. Asking clarifying questions 6. Retelling experiences 7. Ability to concentrate and follow instructions 8. Asking about the meaning of unknown words 9. Laughing/being serious at relevant times during story-telling 10. Participating actively in discussions 11. Understanding abstract language (figurative speech)

1.2.1 Recruitment and population

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 12,491 inhabitants in Langeland municipality in 2020 to 50,676 inhabitants in Hoeje-Taastrup municipality), and the municipalities are distributed across the country, which means that the sample has a diverse demographic composition (see Figure 1.2).

Table 1.2: Examples of text messages within each language domain addressed by the Danish TipsByText curriculum

	Monday: FACT	**Wednesday: TIP**	**Friday: GROWTH**
Vocabulary	Children develop their language skills when they hear new words. Try a new word every day. When your child says shoe, you can answer: tennis shoe.	Talk with your child about getting dressed. Say: you're buttoning your trousers, you're putting on your shirt, you're zipping up your coat...	By using more words and new words, you're boosting your child's language skills. When your child says apple, you can answer: RED apple.
Early literacy	Knowing the letters and their sounds makes learning to read easier. Take one letter at a time to make it easier!	Look for letters on your way home from kindergarten. Start with the first letter of your child's name. Make it a game! Ask: Who can spot it first?	Keep showing your child the first letter of their name - and when you do, say the sound out loud! E.g. the "aaa" sound in Adam.
Communication	Children love knowing lots of things about their family. Looking at family photos is a good way to start a conversation about family.	Talk about the people in your family photos. Tell your child about a fun experience you had with someone in the photo. What did you do?	You can help your child learn by talking about differences such as age or height. Ask: Who's the youngest and oldest in the family photo?

The collaborating municipalities decided which public kindergartens would participate in the RCT; in two out of five collaborating municipalities, all public kindergartens participated, and in three municipalities, 2/3 of public kindergartens participated. In the municipalities where not all public kindergartens participated, the participating kindergartens were typically selected from lower socioeconomic areas, and families from lower socioeconomic strata are thus slightly over-represented in those municipalities. Appendix

Figure 1.2: Map of Denmark with participating municipalities



Table A.1 shows how the sample averages differ from the non-sample averages within the collaborating municipalities.

The population consists of all children born between the years 2013 and 2016 who were enrolled in the participating kindergartens by October 1, 2019. In total, the population

consists of 3,804 children. Table 1.3 shows how the population is distributed across the five collaborating municipalities.

Table 1.3: Overview of the sample across the collaborating municipalities

Municipality	Number of children	Number of parents/ phone numbers	Number of kindergartens
Høje-Taastrup	1,220	1,962	25
Ikast-Brande	1,261	2,023	19
Langeland	210	346	5
Lejre	678	1,148	15
Middelfart	435	717	6
Total	3,804	6,196	70

The RCT included an opt-out design, which means that all children in the population and their parents were enrolled in the RCT by default. Parents received information about the RCT via a pamphlet that was delivered in printed form via the participating kindergartens. [2] The pamphlet describes the intervention, the evaluation (including the randomization), the data collection, the research team, and the opt-out design of the intervention. The material also included a phone number and email address that the parents could contact if they wished to opt out of the trial before the data collection started. Furthermore, parents could opt-out of the intervention at any time after having received the first text message by responding "STOP."

Parental phone numbers were provided to the research team from the municipal administrative records. Thus, the intervention's text messages were delivered to the same phone

[2] To increase the likelihood that the information reached all parents, the pamphlet was also uploaded to most kindergartens' intranets. The information pamphlet was translated into eight languages; all parents received the printed version in Danish, and for each child, the personnel in the kindergarten decided whether to attach a translated version.

numbers the kindergarten pedagogues use if they need to reach the parents, which means that parents have a strong incentive to keep these phone numbers up-to-date.[3]

Children in the population were included in the randomization if three conditions were met on January 1, 2020: (1) the parents of that child had not opted out of the trial, (2) the municipality had provided at least one parental phone number for that child, (3) the kindergarten pedagogues deemed the child fit to participate in the language assessment. In total, **3,611** children were included in the randomization. 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]

1.2.2 Randomization Level and Method

Randomization of treatment status was done at the family level to ensure that everyone within a family received the same treatment. At the time of randomization, we did not have data on which children were siblings, so to minimize spillovers, we categorized all children who had at least one of the registered parental phone numbers in common as siblings. Thus, some stepchildren who shared a parent but did not live together might have been categorized as belonging to the same family. To ensure that all children within a family were assigned the same treatment status, randomization was performed for the youngest child

[3] The research team had no way of verifying the phone numbers, and after sending out the first text message, we did receive a few responses from people who informed us that they were not in the target group (e.g. did not have children). This phone number mismatch can have been a result of either parents having changed phone numbers without informing the kindergarten/municipality, or a result of the kindergarten having made typos when entering the phone numbers in the data that was shared with us.

[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 across the kindergartens.

in the family; we subsequently assigned the older siblings the same treatment status. Each family was sent the TipsByText curriculum aimed at their youngest child in the sample because we expected that if there were to be any differential effects by the age of the child, the largest effects would appear to be among the youngest children.

Families were randomized into three different groups: Treatment group, Control group, and Interview group. The treatment group would receive the TipsByText intervention, the control group would receive treatment as usual, and the interview group would receive the intervention and be interviewed about their experience with the intervention. The interview group was removed from the 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.

Randomization of treatment status was done by a computer using the randtreat command in Stata 15.1. First, 120 families were drawn at random (stratified by municipality, and whether the youngest child's baseline language assessment was above or below the median) to be interviewed about their experience with the intervention.[5] Thereafter, the remaining families were randomized 50/50 into the treatment and control groups stratified by three variables:[6]

- The child's kindergarten (70 strata)
- Whether the youngest child scored above/below the median within their age group in the baseline language assessment, or whether their baseline language assessment

[5] 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 (with 130 children) to this group.

[6] In total there were 840 strata, and misfits were allocated using the global-method in the randtreat command, meaning that misfits were assigned in equal proportions to the treatment and control group.

was missing (3 strata)

- The child's age on November 1, 2020 (4 strata)

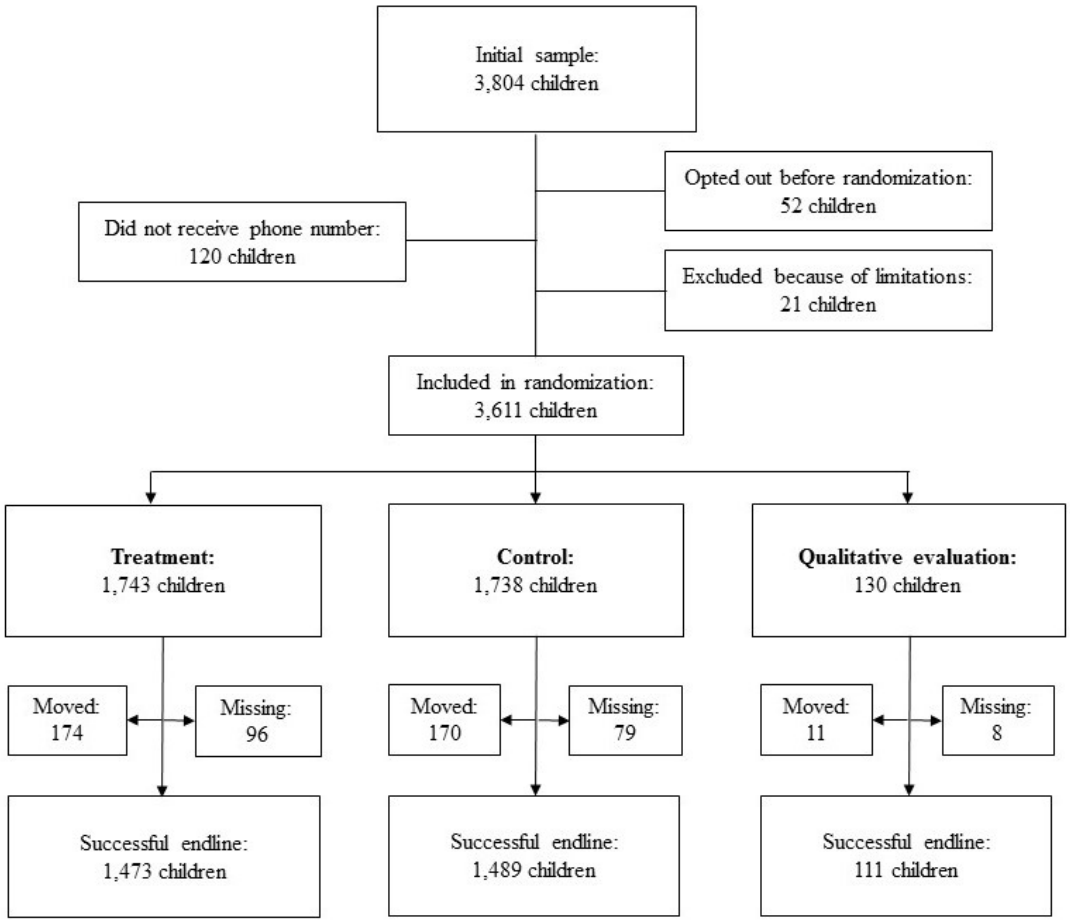
In total, 1,743 children were randomized into the treatment group and 1,738 children were randomized into the control groups (see Figure 1.3 for an overview of the full sample, the randomization sample, and the evaluation sample). To increase statistical power, all analyses will control for the variables that the randomization was stratified on: Kindergarten, baseline language skills within age-group, and age (Bruhn and McKenzie, 2009).

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

1.2.3 Outcome measure: The language assessment, "Sprogvurdering 3-6"

The outcome measure in all the analyses in this chapter will be the children's language skills measured by the Danish language assessment tool, "Sprogvurdering 3-6" (Børne- og Undervisningsministeriet, 2019). The tool is designed to measure the language skills of children in kindergarten (age 3 to 6) and in preschool/the fist year of primary school (ages 5 or 6), and it is widely used in kindergartens/schools as a screening tool to identify which children

Figure 1.3: Overview of sample: From population to final sample



Notes: "Missing" means that the children's endline language assessment is missing despite the child being enrolled at one of the collaborating preschools.

need to be referred to a language specialist or to be monitored for language delays (Bleses et al., 2010).[7] As part of the RCT, the partnering kindergartens increased the scope of their

[7] The tool was developed as part of a federal requirement in Denmark stating that all public schools screen the language skills of children when they enter primary school and all public kindergartens screen the language skills of all two- or three-year-old children who are behind the curve in their language development. In practice, many kindergartens use "Sprogvurdering 3-6" more widely as a preventive tool. For example, in 2018, the year before the RCT, the kindergartens that were part of the RCT assessed the language skills of 81

assessments to cover all children in the RCT in October-December 2019 (baseline) and September-December 2020 (endline).

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/questions. There are four versions targeting kindergarten children at different ages (3 years, 4 years, 5 years, 6 years) and one version targeting school children in their first year of primary school. The older children are assessed in more dimensions of language development than are the younger children (e.g. children are only assessed on their ability to identify letters when they reach age 5), and the difficulty of the questions within each subtest increases with age to avoid ceiling or floor effects (see Table 1.4 for an overview of the subtests included in the language assessment tool, and see Appendix B for more details about the assessment tool and for example questions for each subtest).

The kindergarten versions of the assessment are conducted 1:1 between a pedagogue, who is blind to the child's treatment status, and the child. For most subtests, the pedagogue asks the questions, the child answers, and the pedagogue notes whether the answer was "correct," "incorrect" or "not answered." The subtest "Communication" is directed at the pedagogue; it asks about the child's general use of communicative strategies and the pedagogue fills out whether the child "never," "rarely," "often," or "always" use a specific communication strategy or react in a specific way. The preschool version of the assessment is a class-based written test: The teacher asks the questions out loud and the children note their answers individually. The children who score low on the class-based test are also assessed individually using the kindergarten test for 6-year-old children.

percent of 3-year-olds, 12 percent of 4-year-olds, and 53 percent of 5-year-old children, on average, using the tool "Sprogvurdering 3-6."

Table 1.4: Overview of subtests in the language assessment, "Sprogvurdering 3-6"

Test version Subtest	**Kindergarten: 3 years**	**Kindergarten: 4 years**	**Kindergarten: 5 years**	**Kindergarten: 6 years / Preschool: Individual**	**Preschool: Class-based**
Comprehension	Items 1-20	Items 6-25	Items 11-30	Items 16-35	Items 16-35
Productive vocabulary	Items 1-25	Items 11-35	Items 21-45	Items 27-51	-
Rhyme detection	Items 1-15	Items 1-15	Items 1-15	Items 1-15	Items 1-15
Print-concept knowledge	-	Items 1-16	Items 1-16	Items 1-16	-
Letter identification	-	-	12 letters	12 letters	28 letters
Deletion	-	-	15 words	15 words	-
Auditory discrimination	-	-	-	-	Items 1-20
Communication	Items 1-15	Items 1-15	Items 1-15	-	-

To create the composite measure of language skills, 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 subtest "communication" we calculate the average score on the four-point scale across the 15 questions. Second, we standardize the scores for each subtest because not all subtests use the same scale. This standardization is conducted within each test version (3 years, 4 years, 5 years, 6 years, preschool) because the questions within each subtest differ between test versions. 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 one within each test version (Anderson, 2008).

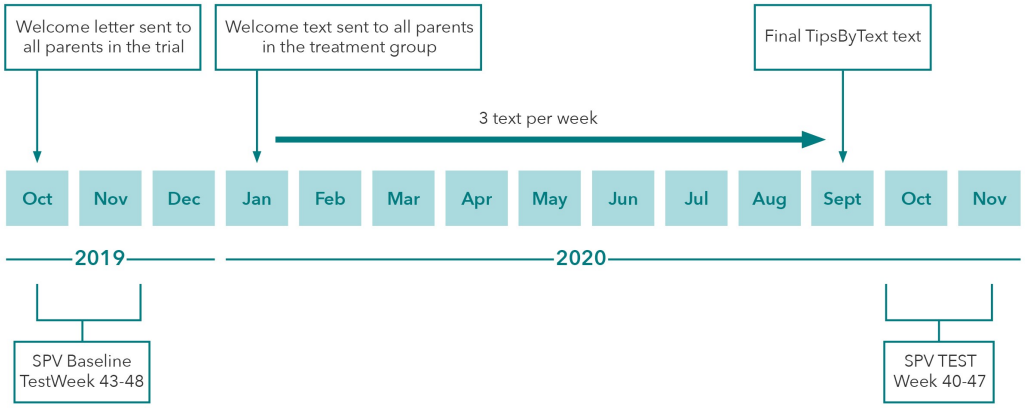
In creating this index, we assume that all measured language dimensions are related

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).

1.2.4 Timeline

The timeline of the RCT and data collection is summarized in Figure 1.4.

Figure 1.4: RCT timeline



1.2.5 Baseline Balance in the Randomized Sample

To test whether there are any important differences between the treatment and control groups despite the randomization, Tables 1.5 and 1.6 compare means of different variables between the two groups for baseline language skills and socioeconomic variables. Appendix Table A.2 shows results for the variables on which the randomization was stratified. Ta-

ble 1.5 shows that there are no important differences in baseline language skills between the treatment and control groups. By chance, the score in the subtest "deletion" is slightly lower for the treatment group than in the control group, and children in the treatment group were assessed at a slightly lower age than children in the control group, but these differences are very small in absolute terms. The difference in "deletion" translates to approximately 13% of a standard deviation, and the difference in age translates to approximately 14 days. Among a wide range of different socioeconomic variables, including parental background, family composition, and child variables, the only significant difference between the treatment and control groups is in the gender composition, where by chance there are slightly more boys in the treatment group than in the control group. We do not expect that the intervention will differentially affect boys and girls beyond through possible differences in baseline scores, which we will control for. Thus, we do not expect this random baseline difference to affect our results. Overall, we conclude that the two groups are similar on average.

1.2.6 Attrition and the Evaluation Sample

A child is defined as attritted if their endline data is missing despite at least one parent being enrolled in the program when the last text message was sent out in September 2020. Because the endline data is collected in the kindergartens, there are two ways a child can attrit (see flow diagram in Figure 1.3). First, a child's endline data will be missing if the child has moved out of the collaborating kindergartens after October 1, 2019 – either by moving to a private kindergarten, to a private school, or to a kindergarten/school outside the five collab-

Table 1.5: Baseline balance: Language skills

	Control mean (SD)	Treatment diff. (SE)	N Treated	N Control
Baseline: Total % correct	0.483	-0.010	1,385	1,400
	(0.197)	(0.007)		
Baseline: Total % wrong	0.433	0.010	1,385	1,400
	(0.171)	(0.006)		
Baseline: Total % missing	0.084	0.000	1,385	1,400
	(0.130)	(0.004)		
Baseline: Comprehension, % correct	0.656	0.000	1,385	1,400
	(0.190)	(0.007)		
Baseline: Vocabulary, % correct	0.417	-0.010	1,385	1,400
	(0.237)	(0.008)		
Baseline: Rhyme detection, % correct	0.472	0.000	1,385	1,400
	(0.356)	(0.011)		
Baseline: Letter identification, % correct	0.472	0.000	856	903
	(0.262)	(0.011)		
Baseline: Deletion, % correct	0.377	-0.040**	406	417
	(0.312)	(0.022)		
Baseline: Print concepts, % correct	0.449	-0.020	406	417
	(0.298)	(0.022)		
Baseline: Communication, mean	2.754	-0.010	1,365	1,375
	(0.655)	(0.022)		
Baseline: Age (years) at language test	4.333	-0.040***	1,385	1,400
	(0.871)	(0.013)		

Notes: This table shows the means of different variables among the control group (column 1), and the results of a linear regression that tests whether the mean of the treatment group is different from the mean of the control group (column 2), clustering standard errors on the level of the family, which is the level of the randomization. Columns 3 and 4 show the number of observations for which data for the given variable is available.

orating municipalities.[8] In this case, a systematic difference in attrition can only happen if the intervention affected whether families decided to move. Secondly, a child's endline data

[8] If a child moves from one participating kindergarten to another participating kindergarten, their endline data will still be collected.

Table 1.6: Baseline balance: Socioeconomic variables

	Control mean (SD)	Treatment diff. (SE)	N Treated	N Control
Indicator: Boy	0.491	0.040**	1743	1738
	(0.500)	(0.017)		
Low birth weight (< 2500 g)	0.056	0.000	1654	1654
	(0.230)	(0.009)		
Number of siblings	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)		
Indicator: Mother working part time	0.183	-0.010	1720	1713
	(0.387)	(0.014)		
Indicator: Father working part time	0.083	0.000	1686	1683
	(0.276)	(0.010)		
Indicator: Father education high school or less	0.633	0.020	1743	1738
	(0.482)	(0.018)		
Indicator: Mother 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 age at birth of child	33.198	-0.280	1685	1684
	(6.024)	(0.225)		
Mother age at birth of child	30.220	-0.040	1690	1685
	(4.981)	(0.184)		

Notes: This table shows the means of different variables among the control group (column 1), and the results of a linear regression that tests whether the mean of the treatment group is different from the mean of the control group (column 2), clustering standard errors on the level of the family, which is the level of the randomization. Columns 3 and 4 show the number of observations for which data for the given variable is available.

will be missing if the kindergarten/preschool did not perform a language assessment for this child despite the child being enrolled. Because the language assessment is conducted by pedagogues/teachers who are uninformed about the children's treatment status, a systematic difference in missing data could only happen if the intervention has large effects on language skills and kindergarten pedagogues are systematically less likely to conduct the language assessment on children who have either stronger or weaker language skills.

A total of 14.36% of children in the sample do not have an endline language assessment; about half of these (7.70% of the full sample) has moved out of the participating kindergartens, and about half (6.66% of the full sample) has a missing endline language assessment despite still being enrolled in one of the participating kindergartens/schools. Importantly, there is no differential attrition across the treatment and control groups: 14.3% of children in the control group and 15.3% of children in the treatment group attrit, and these proportions are not statistically different from each other (see Table 1.7). Thus, it seems that the treatment did not affect the probability to attrit from the sample. This conclusion is confirmed by the fact that the families who attrit in the treatment group are not systematically different from those who attrit in the control group.[9] That is, there is no important differential attrition between the treatment and control groups in baseline language skills and a wide range of demographic variables. All analyses in this chapter will thus be based only on the children for whom we have available endline data – the evaluation sample.

Because (1) there are no baseline differences between the treatment and control groups

[9] There is a small and significant systematic difference in the proportion of boys who attrit, with 45.9% of attrited children in the control group being boys vs. 55.9% of attrited children in the treatment group. However, this difference occurs only among the children who attrited because they moved (see Apprendix Tables A.3 and A.4). As it seems highly unlikely that treatment assignment should affect the probability to move differentially for families with boys vs. girls, we conclude that attrition is independent of treatment.

in the full sample, and (2) there is no differential attrition between the two groups, there are no important systematic baseline differences between the treatment and control group in the evaluation sample (see Appendix Tables A.5 and A.6). However, the evaluation sample is not fully representative of the full sample, parent characteristics predict attrition: Parents with higher education are less likely to attrit, and immigrant parents, parents for whom their first-born child is enrolled in the TipsByText intervention, and parents who are alone are more likely to attrit (see Appendix Table A.7)

1.3 Results

1.3.1 Main effects

The main parameter of interest is the causal treatment effect of the TipsByText intervention on children's language skills. This is the intent-to-treat parameter, which comes from comparing the language outcomes of all children who were randomized into the treatment group to the language outcomes of all children who were randomized into the control group. To estimate the causal effect of the intervention, we run a multivariate linear regression with a composite measure of language skills as the dependent variable and a dummy for whether the child's family was randomized into the treatment group as the main independent variable.

The regression specification is as follows:[10]

$$y_{i,1} = \alpha + \beta T_i + \theta y_{i,0} + \rho C_i + v_i + k_i + \varepsilon_i, \tag{1.1}$$

[10] This empirical specification as well as all following regression analyses in this chapter were pre-specified in AEA's trial registry: http://www.socialscienceregistry.org/trials/6393.

Table 1.7: Differential attrition between the control and treatment groups

	Control mean (SD)	Treatment diff. (SE)	N Treated	N Control
Analysis on full sample				
Attrited	0.143	0.010	1743	1738
	(0.350)	(0.013)		
Moved out of municipality	0.098	0.000	1743	1738
	(0.297)	(0.011)		
Missing language assessment	0.045	0.010	1743	1738
	(0.208)	(0.008)		
Analysis on attrited sample				
Baseline: Total % correct	0.497	0.030	214	206
	(0.211)	(0.021)		
Baseline: Total % wrong	0.428	-0.020	214	206
	(0.181)	(0.018)		
Baseline: Total % missing	0.075	-0.010	214	206
	(0.140)	(0.013)		
Indicator: Boy	0.459	0.100**	256	244
	(0.499)	(0.045)		
Indicator: Firstborn child (mother)	0.480	0.030	256	244
	(0.501)	(0.044)		
Indicator: One adult in home	0.154	0.000	249	241
	(0.361)	(0.034)		
Indicator: First/second generation immigrant	0.287	-0.030	256	244
	(0.453)	(0.041)		
Indicator: Dad's education high school or less	0.611	-0.020	256	244
	(0.489)	(0.046)		
Indicator: Mom's education high school or less	0.475	-0.010	256	244
	(0.500)	(0.047)		

Notes: This table shows the means of different variables among the control group (column 1) and the results of a linear regression that tests whether the mean of the treatment group is different from the mean of the control group (column 2), clustering standard errors on the level of the family, which is the level of the randomization. Columns 3 and 4 show the number of observations for which data for the given variable is available. Comparisons in the top panel are based on the full sample, and comparisons in the bottom panel are based only on the group of children who was missing an endline language assessment.

where $y_{i,1}$ is the composite measure of language skills for individual i at endline standardized within age-group. T_i is the treatment dummy equal to 1 for the treatment group and 0 for the control group. $y_{i,0}$ is the composite measure of language skills for individual i at baseline. C_i is a vector of controls, including a range of demographic and income variables, identified to be correlated with early language development through the machine learning method LASSO (least absolute shrinkage and selection operator) (Ludwig et al., 2019).

The regression controls for kindergarten fixed effects, k, and language assessment version fixed effect, v, such that children in the treatment group are only compared to children in the control group within the same kindergarten and within the same test version. Kindergarten fixed effects are included because the language assessment is conducted in the kindergartens, and thus kindergarten fixed effects control for possible surveyor fixed effects. As kindergartens are nested within the five demographically different municipalities, controlling for kindergarten fixed effects also controls for municipality fixed effects. Test version fixed effects are included to ensure that children in the treatment group are only compared to children in the control group who answered the same set of questions in the language assessment. We will use robust standard errors clustered on the family level, as the randomization is carried out at the family level, whereas the language outcomes are measured at the level of the individual child (Abadie et al., 2017).[11]

As the outcome measure, y_i, is a standardized z-score, the point estimate, $\hat{\beta}$, on the intent-

[11] We pre-specified that the control vector, C_i would include variables on which the randomization was stratified, as this often increases power (Bruhn and McKenzie, 2009). The randomization was stratified on 1. kindergarten at baseline, 2. age as a proxy for test version at endline, and 3. whether the child scored above or below the median in the baseline language assessment. We control for the version of each of these variables that makes the most theoretical sense; i.e., the regression controls for 1. kindergarten at *endline*, 2. *realized* test version at endline, and 3. the *continuous*, standardized baseline measure. Robustness tables in the Appendix show that the results are robust to controlling for the actual stratification variables.

to-treat treatment dummy, T_i, from this multivariate regression constitutes our best estimate of the standardized within-sample effect size. The results of four versions of the empirical specification in equation 1.1 are presented in Table 1.8. Column (1) includes no controls, column (2) includes only the baseline language measure as control, column (3) includes the baseline language control but does not control for kindergarten fixed effects because doing so excludes singletons, and column (4), our main empirical specification, includes the full set of controls.

Table 1.8: Treatment effect of the TipsByText intervention on children's language skills

	Dependent variable: Standardized index of language skills			
	(1)	(2)	(3)	(4)
Treatment	-0.041	-0.028	-0.029	-0.030
	(0.035)	(0.026)	(0.027)	(0.025)
Kindergarten FE	✓	✓		✓
Test version FE	✓	✓	✓	✓
Baseline controls		✓	✓	✓
LASSO controls				✓
SE cluster level	Family	Family	Family	Family
Observations	2,977	2,977	2,981	2,981
R^2	0.206	0.554	0.479	0,562

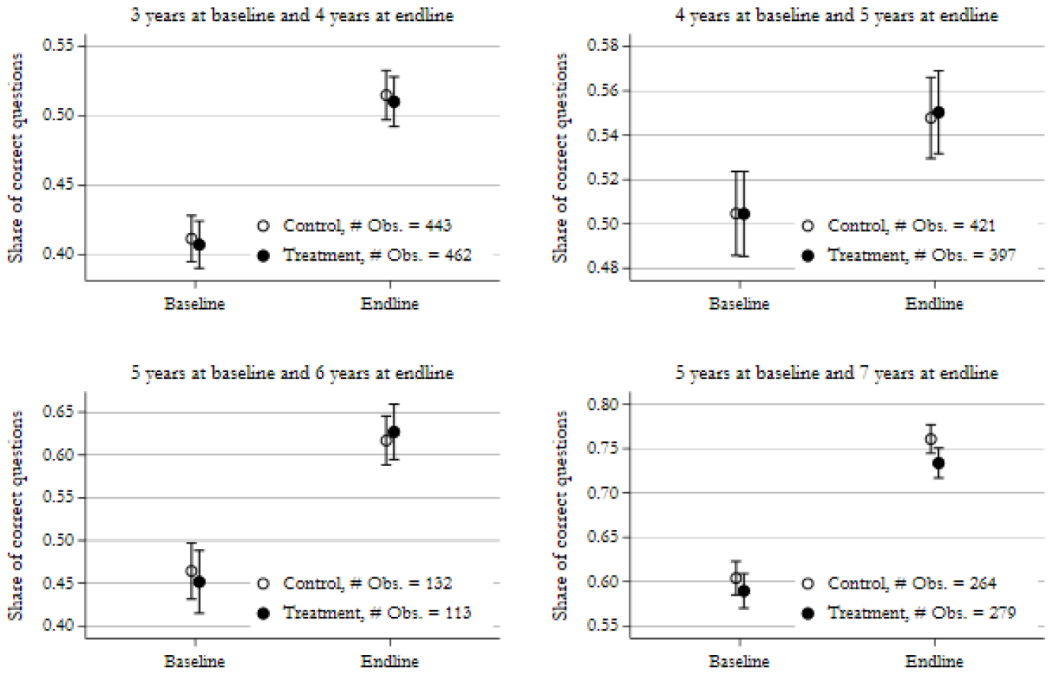
Notes: The table presents the results of a linear regression of the standardized language index on a treatment indicator, the standardized baseline measure, an indicator for whether baseline was missing, and a set of controls selected by LASSO. The result of the main specification, which was pre-specified, is indicated in blue. Kindergarten singletons are excluded in columns (1) and (2). * p < 0.10, ** p < 0.05, *** p < 0.01.

As shown in Table 1.8, the point estimate on the treatment dummy is extremely close to zero regardless of the details of the empirical specification; the difference in language skills

between the treatment and control groups is between −0.025 S.D. and −0.038 S.D. Looking at the standard errors, we see that these null effects are very tightly estimated, indicating that the null effect is unlikely to be a result of noise or an under-powered negative effect.

The lack of a positive effect is also evident when looking at other variants of the outcomes measure, such as the proportion of correctly answered questions, or the change in the proportion of correctly answered questions between the baseline and endline language assessment (see Figure 1.5 or Appendix Table A.8). These results suggest that the null result is not due to how the index is constructed.

Figure 1.5: Graphical representation of the treatment effect on children's language skills



Notes: The graphs present the share of correctly answered questions in the language assessment across treatment status, and across baseline and endline measures. Each of the four graphs presents a unique combination of test versions used to assess the child at baseline and endline. Vertical lines represent confidence bands at a 95%-level.

Based on these regression results, we conclude that the TipsByText intervention did not improve average language skills for kindergarten children as measured by the language assessment, "Sprogvurdering 3-6." Next, we explore whether the intervention improved any dimension of language skills and whether any subgroups of children experienced a positive treatment effect.

1.3.2 Treatment effects on individual language dimensions

The regression results presented in Table 1.8 showed the treatment effects on the standardized index of language skills. As explained in Section 1.2.3, this index is calculated as the average standardized scores across 4–7 subtests depending on the age of the child. The index assumes 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. If this assumption is not satisfied, the null effect on this index could be a result of a significant *positive* effect in some subtests and a significant *negative* effect in some subtests.

The treatment effects for each subtest that enters into the index are shown in Table 1.9. These effects are estimated using the same specification as in Equation 1.1, except we only control for the baseline variable to maximize degrees of freedom, as Table 1.8 showed that the treatment effect was essentially unaffected by including the full set of controls (compare columns 2 and 4).

The treatment effect is null or very close to null in each language dimension (see Table 1.9). The effect on "Comprehension" is negative and statistically significant on a 10% significant level, but the point estimate is still very small (-0.053 S.D.). These additional

null effects indicate that the overall null effect is not driven by the index averaging over some positive and some negative effects. Interestingly, the signs of the point estimates of the treatment effect do vary between the eight dimensions: three coefficients are positive, whereas five coefficients are negative. We interpret this variation across subtests as evidence that the negative sign of the main effect in Table 1.8 is not the result of an overall negative effect of the intervention.

Table 1.9: Treatment effect of the TipsByText intervention on different dimensions of language skills

	(1) Comprehension	(2) Vocabulary	(3) Rhyme detection	(4) Concepts of print
Treatment	-0.053*	0.007	0.006	-0.009
	(0.030)	(0.028)	(0.031)	(0.032)
Kindergarten FE	✓	✓	✓	✓
Test version FE	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓
SE cluster level	Family	Family	Family	Family
Observations	2,977	2,316	2,977	2,211
R^2	0.412	0.586	0.346	0.422
	(5) Letter identification	(6) Deletion	(7) Auditory discrimination	(8) Communication
Treatment	-0.021	-0.002	-0.104	0.024
	(0.042)	(0.048)	(0.064)	(0.034)
Kindergarten FE	✓	✓	✓	✓
Test version FE	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓
SE cluster level	Family	Family	Family	Family
Observations	1,830	1,169	659	2,031
R^2	0.324	0.400	0.392	0.435

Notes: The table presents the results of a linear regression of the standardized score in each subtest in the language assessment on a treatment indicator, the standardized baseline measure, and an indicator for whether baseline was missing. The number of observations differ between columns because children of different ages answer different subtests. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

1.3.3 Comparison of treatment and control beyond the mean

Tables 1.8 and 1.9 showed that the *average* treatment effect of the intervention is zero. However, many different distributions can lead to the same mean, and an average treatment effect of zero might conceal that the treatment affected the *distribution* of the outcome even if it did not affect the mean.

To study whether the TipsByText treatment affected the distribution of the outcome measure, the full distribution of the standardized endline language scores of the treatment and control groups are plotted in Figure 1.6. The distributions of the endline language scores are almost identical in the treatment and control groups, further cementing the evidence that the average null effect seems to come from a true lack of effect of the intervention on the children's language skills.[12] The bell-shaped distribution of outcomes also indicates that the lack of positive effect is not the result of a ceiling effect in the language assessment.

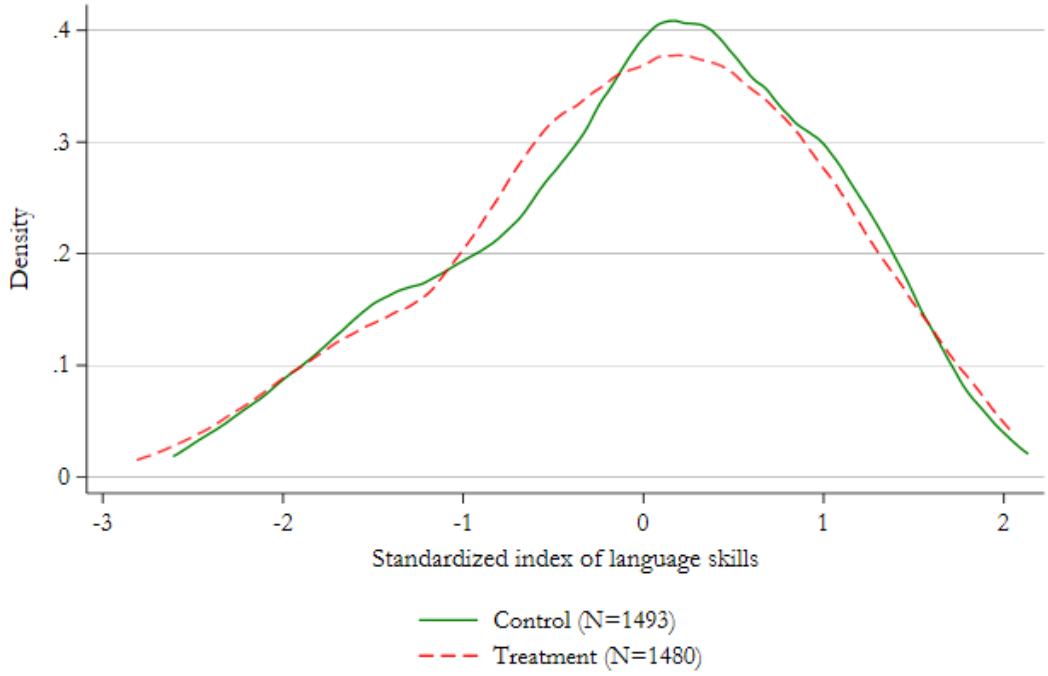
Overall, we conclude that the TipsByText intervention did not improve children's language skills as measured by, "Sprogvurdering 3–6." In the next section, we explore whether there are differential treatment effects across subgroups.

1.3.4 Differential Treatment Effects

Due to the positive correlation between parents' child-directed speech and children's language skills, we expected TipsByText to have a larger effect among children of parents who

[12] There does seem to be slightly more mass between the values -1 and 0 in the treatment group than in the control group, but regression analyses in Appendix Table A.9 show that there are no statistically significant differences in the probability between the two groups of being in being in the first, second, third, or fourth quartile of the outcome distribution within a given age group between the treatment and control group.

Figure 1.6: Distribution of language skills in the treatment vs. control group



Notes: The figure presents 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 a Gaussian kernel function.

engaged relatively less in active language stimulation with their children before the intervention (*Hypothesis 1.2*). We, therefore, expected the intervention to decrease the language gaps among kindergarten children.

Ideally, we would be able to calculate differential treatment effects based on the quality and quantity of parents' child-directed speech at baseline. Unfortunately, we do not have direct measures of the children's home learning environments. Instead, we use the baseline language score as a crude proxy for the home learning environment. *Hypothesis 1.2* thus translates into an expectation that the intervention will have relatively larger effects for the

children who score relatively lower in the baseline language assessment. Furthermore, because research has shown that the quality and quantity of child-directed speech is correlated with socioeconomic status (SES, e.g. Rowe, 2008), we expect that the intervention will have relatively larger effects for children of parents with relatively lower educations according to (according to Hoff, 2013, maternal education is the most relevant component of SES for children's language development). Finally, as one of the strongest demographic predictors of children's baseline language skills in our data is whether the child is a first- or second-generation immigrant, we also expect immigrants and children of immigrants to benefit relatively more from the intervention.

We test *Hypothesis 1.2* using the same multivariate linear regression framework from Equation 1.1, but in four separate regressions we will add a term that interacts the treatment dummy with a dummy for (a) whether the child scored below the median in the baseline language assessment, (b) whether the mother has a high school degree or less, (c) whether the father has a high school degree or less, or (d) whether both parents are immigrants.

The regression specification is as follows:

$$y_{i,1} = \alpha + \beta T_i + \gamma I_i + \delta I_i \times T_i + \theta y_{i,0} + + v_i + k_i + \varepsilon_i, \tag{1.2}$$

where y_i again is the composite measure of language skills for individual i, T_i is the treatment dummy equal to 1 for the treatment group and 0 for the control group, I is the subgroup dummy equal to 1 if e.g. the child scored below the median at baseline, v captures test version fixed effects, k captures kindergarten fixed effects, and ε_i is an error term. Standard errors are again clustered at the level of the family, and we control only for baseline

language scores to maximize degrees of freedom.

The coefficients of interest in these regressions are the interaction terms, δ, which show the treatment effects of group I (e.g. children who score *below* the median at baseline) over and above the treatment effect of the rest of the children (e.g. children who score *above* the median at baseline). The coefficients, δ, thus show whether the TipsByText intervention deferentially affected the groups of children who we expected would experience the largest benefits of the intervention. Results are shown in Table 1.10; Appendix Table A.10 show regression results separately for the four subgroups of interest.

As shown in Table 1.10 column (1), there are no differential treatment effects based on whether the child scored above or below the median in the baseline language assessment: The children in the treatment group who scored below the median in the baseline language assessment, scored on average 0.026 S.D. higher at endline than did the children in the treatment group who scored above the median at baseline. Although the point estimate difference is in the expected direction, with a higher estimated effect among the children who scored above the median at baseline, this difference is very small and not statistically significant.

The same pattern of no differential treatment effects among any groups appears when we look at the three other groups. The point estimate of the treatment effect for the children of fathers with high school as the highest education (column 3) is positive but insignificant and extremely small, 0.026 S.D. The same is true for the point estimate of the differential treatment effect for the children of immigrants (column 4) which is 0.021 S.D. In the opposite direction of our hypothesis, the point estimate of the treatment effect for the children of mothers with high school as the highest education (column 3) is actually negative,

Table 1.10: Differential treatment effects of the TipsByText intervention

	Dependent variable: Standardized index of language skills			
	(1)	(2)	(3)	(4)
Treatment	-0.029	0.000	-0.042	-0.032
	(0.033)	(0.038)	(0.043)	(0.028)
Below median at baseline test	-0.046			
	(0.048)			
Below median at baseline test × Treatment	0.026			
	(0.050)			
Mother's education high school or less		-0.110***		
		(0.037)		
Mothers education high school or less × Treatment		-0.052		
		(0.053)		
Father's education high school or less			-0.130***	
			(0.038)	
Fathers education high school or less × Treatment			0.026	
			(0.055)	
First/second generation immigrant				-0.222***
				(0.058)
First/second generation immigrant × Treatment				0.021
				(0.070)
Kindergarten FE	✓	✓	✓	✓
Test version FE	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓
SE cluster level	Family	Family	Family	Family
Observations	2,780	2,977	2,977	2,977
R^2	0.587	0.558	0.556	0.558

Notes: The table presents the results of a regression of the standardized language on a treatment indicator, a subgroup group dummy, an interaction of the treatment indicator and the subgroup dummy, the standardized baseline measure, and an indicator for whether baseline was missing. The number of observations differ between groups because information on parental education and immigration status is missing for some children. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

−0.052 S.D., indicating that, if anything, the effect was *lower* among this group.

We thus conclude that the TipsByText intervention was not successful in decreasing gaps in children's early language skills, as measured by "Sprogvurdering 3–6." Nor was it successful in improving language skills for any of the pre-specified subgroups.

1.3.5 What the estimated null effect can tell us about the true effect of the TipsByText intervention

In this section, we discuss what the observed null results allow us to conclude about the effect of the intervention given (1) the ex post minimum detectable effect size, and (2) the negative predictive value.

Ex post minimum detectable effect size

The sample of this study is determined by power calculations for a minimum detectable effect of MDE=0.1 S.D. with 80% power, a false positive rate of $\alpha = 0.05$, and an assumed $R^2 = 0.4$, and attrition=20% (see figure A.1 for the power simulations upon which the sample was determined).

These power calculations are based on assumptions made before we received the realized data. We can revisit these power calculations with the realized data and calculate the *ex post* MDE using the observed standard errors from the main analysis presented in Table 1.8. To have 80% power, the true value of the parameter must be 2.8 standard errors from the comparison point, which in this study is zero. That is, for a true value of a normally distributed parameter equal to 2.8∗S.E., 80% of the mass will be to the left of the critical value of 1.96 (Gelman and Hill, 2007, Chapter 20). The ex post MDE with 80% power and a false posi-

tive rate of $\alpha = 0.05$ is thus given by MDE=2.8∗S.E.

In this study, the S.E. for the main analysis in Table 1.8 is 0.025. Thus, the ex post MDE is $2.8 * 0.025 = 0.07$ S.D., which is even smaller than the ex ante MDE of 0.1 S.D. The interpretation of the ex post MDE of 0.07 S.D. is that with the realized variance of the data and the realized explanatory power of the control variables, we should be been able to detect effects down to 0.07 S.D. We interpret the fact that we observe null effects despite a fairly small ex post MDE to be a strong indicator that lack of power cannot explain the lack of a significant positive effect of the TipsByText intervention on children's language skills.

Negative predictive value

In Bayesian statistics, the *negative predictive value* is the probability that a measured null effect reflects a true null effect (or in this case, an effect smaller than our minimum detectable effect, MDE of 0.1 S.D.). This value depends on the sensitivity, specificity, and prevalence:

$$NPV = \frac{specificity \times (1 - prevalence)}{specificity \times (1 - prevalence) + (1 - sensitivity) \times prevalence}, \tag{1.3}$$

where $specificity = (1 - \alpha)$ is the rate of true negatives (i.e. the probability of correctly failing to reject a null hypothesis when the null hypothesis is false), $sensitivity = (1 - \kappa)$ is the rate of true positives (i.e. the *power* of the test of an effect greater than the MDE), and $prevalence = \pi$ is the *prior probability* of observing an effect greater than the MDE.

We know from our power calculations that for an MDE of 0.1 S.D., the rate of false positives is $\alpha = 0.05$ and the power is $(1 - \kappa) = 0.80$. Thus, Equation 1.3 can be written as:

$$NPV = \frac{(1-\alpha) \times (1-\pi)}{(1-\alpha) \times (1-\pi) + \kappa \times \pi} = \frac{0.95 \times (1-\pi)}{0.95 \times (1-\pi) + 0.02 \times \pi} \tag{1.4}$$

The interpretation of the NPV is: Given that we observe no effect of the intervention, NPV is the probability that there actually is no effect. As boundary examples, if we are ex ante *certain* that the intervention does have an effect greater than the MDE, $\pi = 1$ and $NPV = 0$. That is if we are ex ante *certain* that the intervention does have an effect greater than MDE, the posterior probability that the intervention has no effect given that we observe no effect is 0%. On the contrary, if we are certain that the intervention has *no* effect, $\pi = 0$, and $NPV = 1$. That is the posterior probability that the intervention has no effect given that we observe no effect is 100%. Thus, if we are *certain* that the intervention either has a positive effect or no effect, no experimental result will make us update our prior.

If, instead, we are agnostic about whether the intervention had an effect, such that $\pi = 0.5$, we get that $NPV = 0.83$. That is, given the power of the test of $\beta > MDE$ and the false positive rate, the posterior probability of $\beta > 0.1$ S.D. is 83% if the prior probability is 50%. If, instead, we expected that the population parameter $\beta > 0.1$ S.D. with 80% probability, the posterior probability is 54%. Or, if we were fairly certain that the intervention does in fact have a positive effect greater than the MDE, the current experiment would make us agnostic about whether the intervention has an effect.

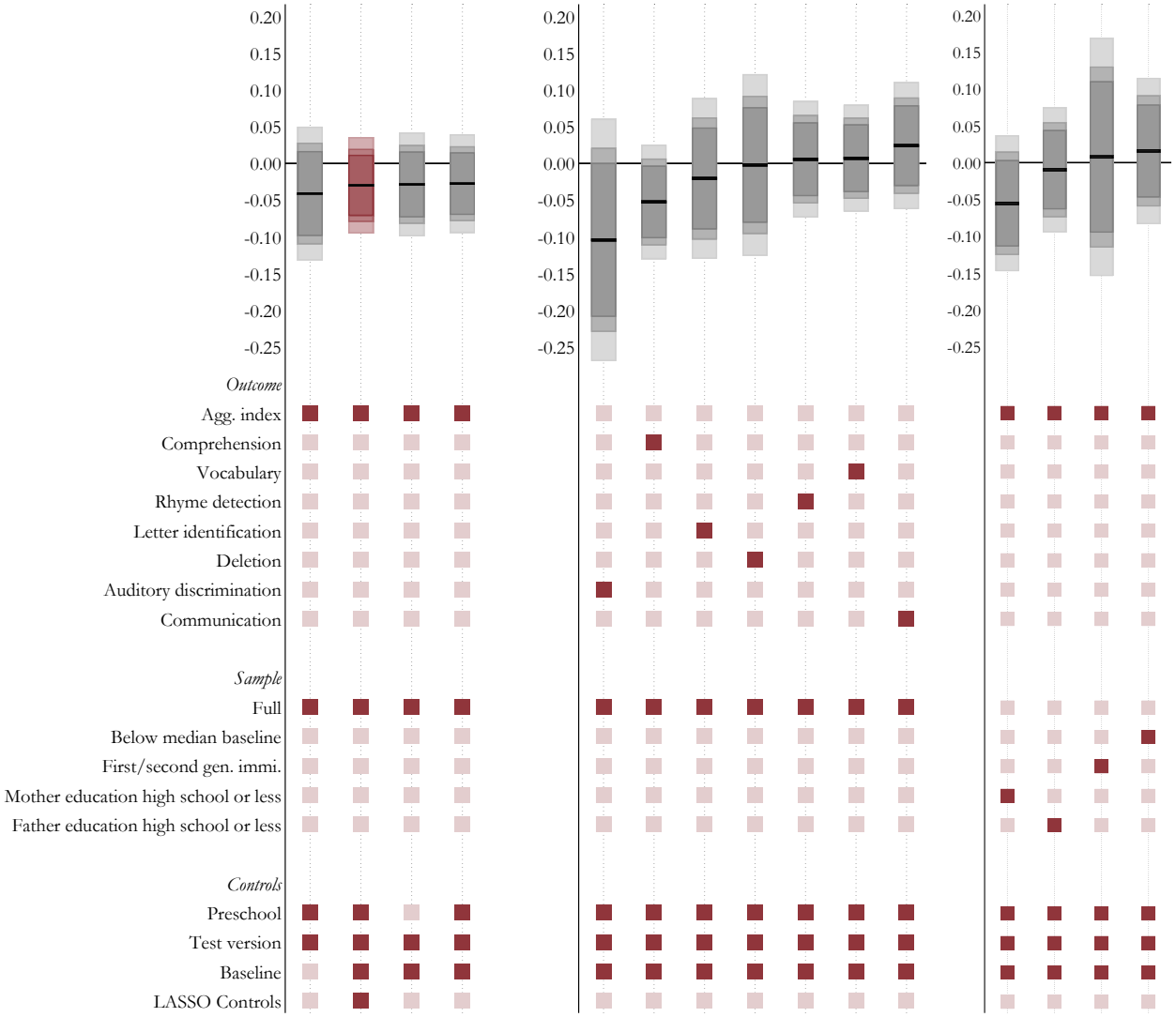
1.4 Conclusion

In this chapter, we show that the TipsByText intervention was not successful in improving the language skills among children in the current sample (Figure 1.7 summarizes all results presented in Chapter 1 graphically). We show that the intervention did not affect average

language skills in the treatment group, that it did not affect any of the dimensions of language skills measured by "Sprogvurdering 3-6," and that it did not improve the language skills of any of the groups of children who we had expected would benefit the most from the intervention. As the null effects are very tightly estimated and as the ex post minimum detectable effect size is small (0.07 S.D.), we conclude that lack of positive effect of the intervention is not likely to be driven by lack of power, but instead that it is likely the result of a true null effect of the intervention on children's language skills.

In Chapter 2, we will use multiple available data sources to try to answer the question: "Why was the TipsByText intervention not successful in improving the language skills among children in the current sample."

Figure 1.7: Overview of effect sizes in Chapter 1 for different empirical specifications



Notes: The figure shows the coefficients and confidence intervals for different regression specifications of the effect of the TipsByText intervention on children's language skills. Coefficients are depicted by black horizontal lines. The vertical bars, from lightest to darkest, denote the 90%, 95% and 99% confidence intervals, respectively. The red bar indicates our main specification. The bottom panel indicates the regression specification in terms of the outcome measure, the sample, and the set of control variables.

"It's a good question whether it has made a difference for my child. I don't necessarily think so. But it's made a difference to ME."

– Birgitte, recipient of the TipsByText intervention

2

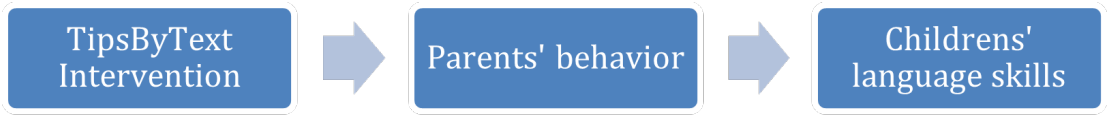
Why did TipsByText not improve children's language skills?

Chapter 1 established that the TipsByText intervention did not improve children's language skills, but it left open the question of "why not?" This chapter will explore this

question using multiple additional data sources.

The effect of the TipsByText intervention relies on the simplified theory of change presented in Figure 2.1. There are thus 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. In this chapter, we will explore each of these two possible explanations in turn.[1]

Figure 2.1: Theory of change for the TipsByText Intervention



2.1 Exploring the link between the intervention and parent behavior

In the original impact evaluation of Ready4K!, the American version of TipsBytext, York, Loeb and Doss (2019) found that the intervention increased parents' engagement in home literacy practices compared to the control group (the effect was 0.156 S.D. on a composite variable of home literacy practices, significant on a 10% significance-level across the two cohorts). The effect was primarily driven by positive effects on book-related behavior, such as pointing to different parts of the book or looking at pictures in books, and reciting nursery rhymes, but the point estimates were positive on almost all measured behaviors, including pointing out letters and words in the environment, introducing new words using household objects, and telling children stories. While these analyses are slightly under-powered ($N = 536$) and rely on self-reported behaviors, the results are consistent with the Tips-

[1] None of the analyses presented in this chapter have been pre-registered.

ByText intervention having been successful in affecting parent behavior in a different context/sample.

Unfortunately, we do not have data on parental behavior in our complete sample. Instead, we will set up and test three necessary – but not sufficient – conditions for the intervention to have affected parent behavior. We will also draw on qualitative and non-representative survey data to ask whether parent behavior was affected for some parents.

The three necessary – but not sufficient – conditions for the intervention to have affected parental behavior are as follows:

1. *Successful delivery*: The text messages were successfully delivered to the parents
2. *Compliance*: Parents did not opt out of the intervention
3. *Fidelity*: Parents read and used the content of the text messages

In this section, we will use different available data sources to determine whether either of these conditions were *not* met. We show data that are consistent with each of the conditions being met successfully, indicating that the intervention did, in fact, affect parental behavior – at least for some parents.

2.1.1 Were the text messages not successfully delivered to the parents?

The first necessary condition for the intervention to have affected parental behavior is that the intervention's text messages were successfully delivered to the parents. There are two possible challenges to this condition having been met. First, all texts were uploaded to and sent from an electronic platform which communicated automatically with a text message service to send out the text messages. If either of these two electronic services failed or if

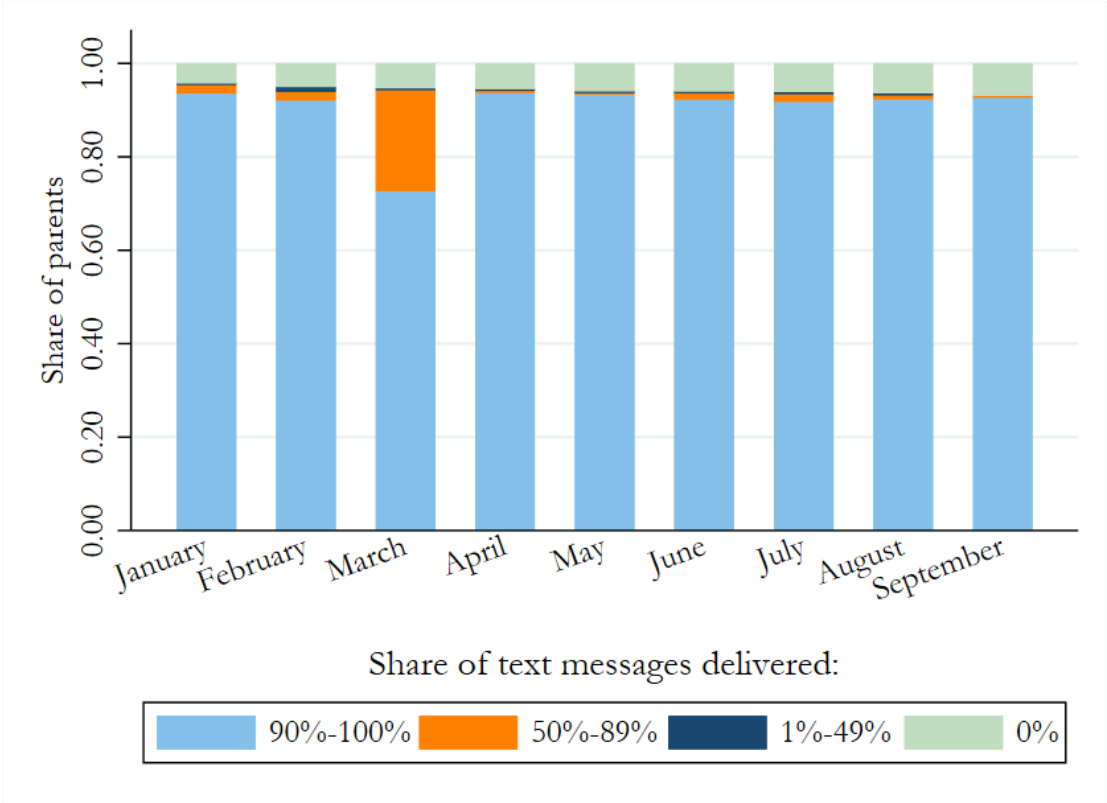
communication between the services failed at critical times, the text messages would not be sent out. Second, parents were enrolled in the intervention with a specific phone number, and to minimize spillovers into the control group, we did not allow parents to switch phone numbers during the trial. Thus, if a parent changed phone numbers during the eight months of the intervention, they would no longer receive the texts.

We test whether parents received the intervention's text messages by using data on delivery notifications from the electronic platform. We find that more than 90% of the parents who did not opt our of the intervention received more than 90% of the text messages each month, except for in March, where there was a technical issue that resulted in three text messages not being delivered (see Figure 2.2). Thus, lack of delivery of the intervention's text messages to the parents cannot explain the lack of positive effect of the intervention on children's language development.

2.1.2 Did too many parents opt out of the intervention?

The second necessary condition for the intervention to have affected parental behavior is that the parents stayed enrolled in the intervention. Parents could opt out of the intervention at any point after the first text was received, by responding "STOP" to the number from which the intervention was send. The results in Chapter 1 rely on treatment-on-the-treated analyses; that is, analyses in which *all* children in the treatment group are compared to *all* children in the control group regardless of whether the parents of the children in the treatment group complied with the intervention or not. Thus, one reason for the lack of a positive effect on children's language skills could be that too many parents opted out of the intervention and thus that the intent-to-treat estimate includes too many children whose

Figure 2.2: The proportion of parents who received a given proportion of the text messages in a given month.



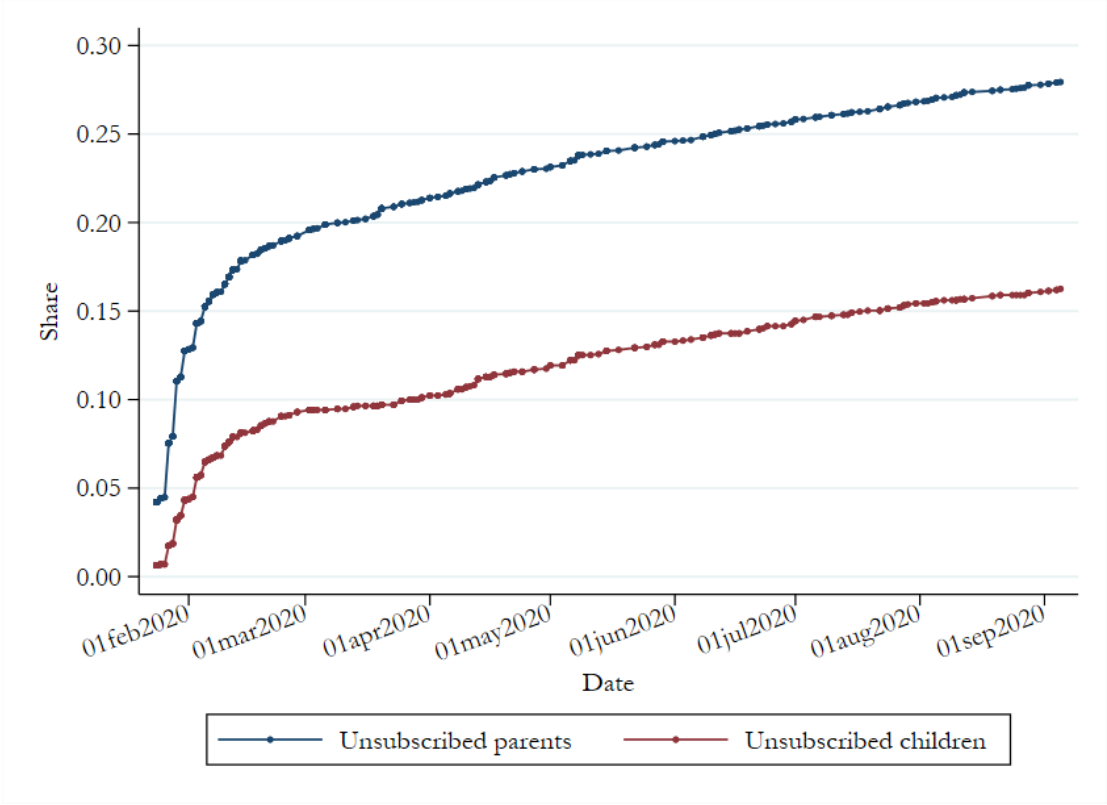
Notes: The figure shows the proportion of parents who have received a given proportion of text messages. The figure includes both parents who opted out at some point during the intervention, and parents who didn't. The proportion of received messages for parents who opted out in the middle of a month are calculated as the proportion of messages received out of the text messages they should have received before opting out.

parents actually did not receive the intervention.

We test whether parents opted out of the intervention by looking at how many parents responded "STOP" to the number from which the intervention's text messages were sent. We find that 83.7% of children still had *at least* one parent enrolled in the intervention when the last text message was sent out in September 2020, and 60.4% of children still

had all parents enrolled in the intervention at that point (see Figure 2.3). Accordingly, the treatment-on-the-treated effect, in which the endogenous variable "compliance" is instrumented by the exogenous variable "treatment assignment," is also negative and very close to the intent-to-treat estimate (see Appendix Table A.11). A high opt-out rate is therefore unlikely to explain for the null effect.

Figure 2.3: The proportion of parents who opted out of the intervention.



Notes: The figure shows the share of parents who opted out of the TipsByText intervention after the first text message was sent out on January 25, 2020, and before the last messages was sent out on September 4, 2020. The share of unsubscribed parents is calculated as the share of parents who unsubscribed out of the total number of parents who was enrolled in the intervention on January 25, 2020. The share of unsubscribed children is calculated as the share of children where all registered parents opted out of the intervention out of the total number of children who was enrolled in the intervention on January 25, 2020.

A further confirmation of this conclusion is that the endline language skills of children in the treatment group do not seem to correlate with the number of text messages the parents received (see Figure 2.4. If anything, by looking only at the bottom panel it looks like the children of parents who stayed enrolled for the whole trial actually scored lower on the endline language test than did children of parents who opted out. However, the top panel indicates that this same trend is present in the baseline test scores, indicating that the trend is driven by differences in characteristics of who opted out, rather than effects of the intervention. This interpretation is confirmed in Appendix table A.12, which shows that parents of children who score higher on the baseline test were more likely to opt out of the intervention.

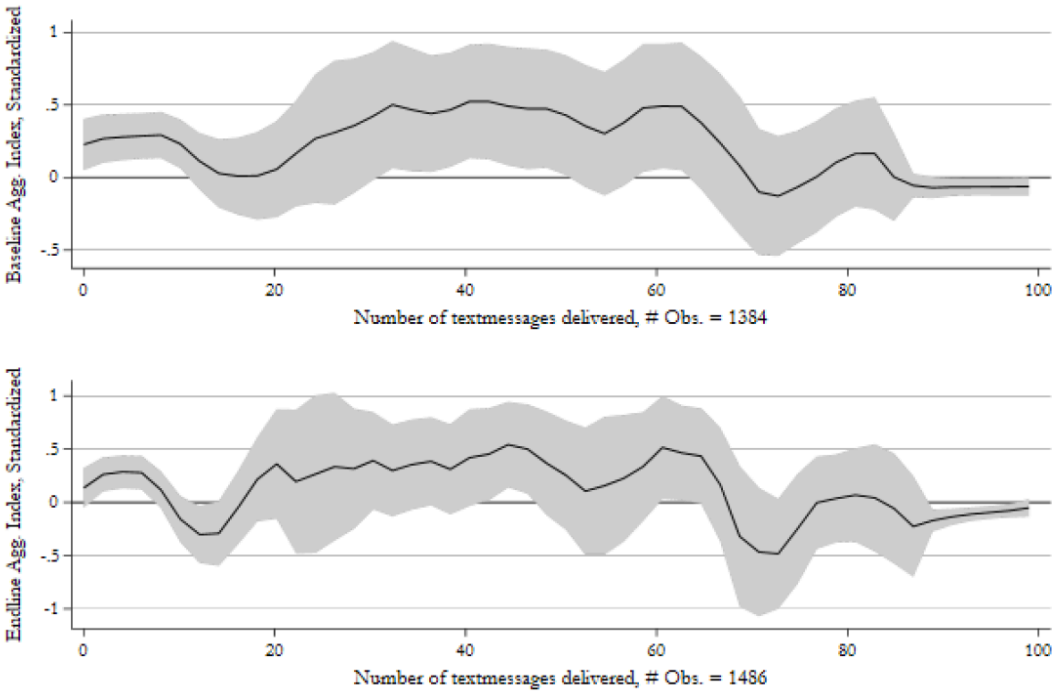
2.1.3 Did parents not read or use the text messages?

The third necessary condition for the intervention to have affected parental behavior is that the parents complied with the intervention; that is, that they read the intervention's tips and implemented them with their children. In this section we characterize parental usage of the intervention using data from a parent survey of a non-representative sample and from a qualitative evaluation that was conducted concurrently with the RCT by an independent team.

Parent survey responses

After the intervention had ended, we sent a short survey via text message to all parents in the treatment group who were still enrolled in the intervention in September (N=2,085). The survey asked parents up to 7 questions about how often they read and used the inter-

Figure 2.4: Dose response of the TipsByText intervention



Notes: The figure shows the average standardized baseline (top panel) and endline (bottom panel) language outcome and 95% confidence intervals for children whose parents have received a given number of text messages. Parents received less than 100% of the text messages either if they opted out or if the text messages failed to be delivered. The number of received text messages for a given child is based on the parent who received the highest number.

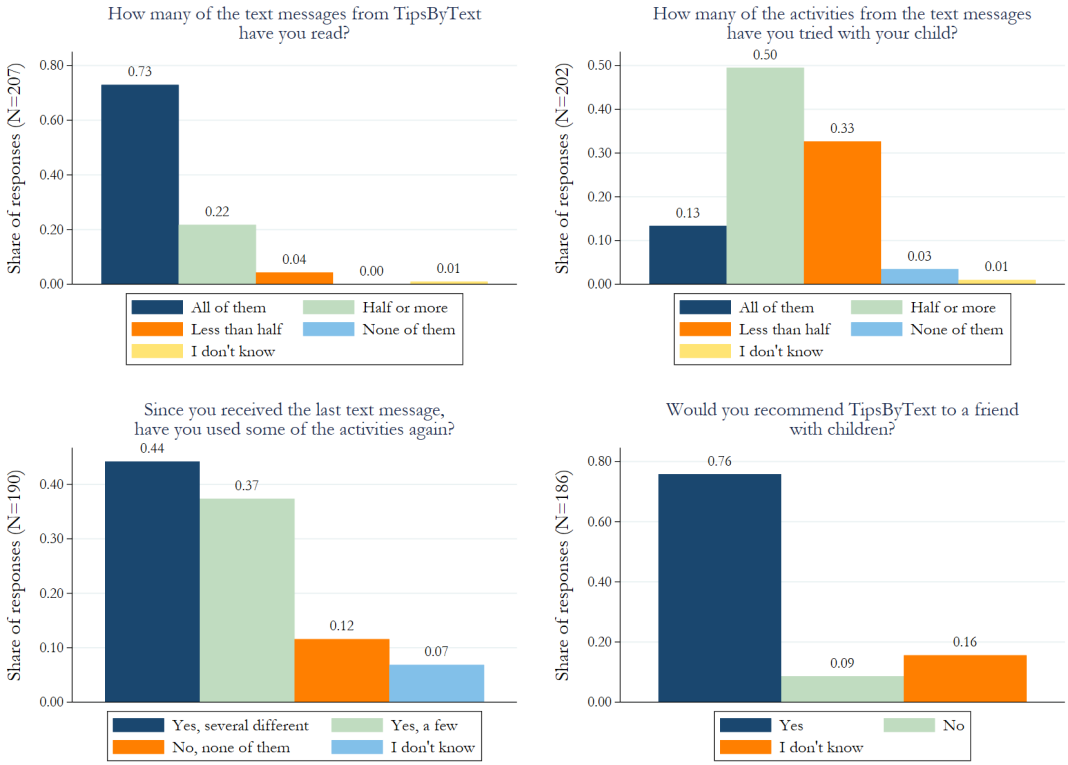
vention's text messages. The survey could be answered either via text message or online. In total, 223 parents, or at least one parent for 10.7% of children, responded to the survey's first substantive question, and 194 parents, or at least one parent for 9.4% of children, responded to all questions. This is clearly a very low response rate, and we acknowledge that the sample of parents who responded is likely to be among the parents who engaged most with the intervention.[2] Therefore, we will not make claims about the entire sample from this selective sample. Instead, we will use the results from this selective group as "proof of concept" of whether some parents engaged with the intervention.

In the parent survey, 95% stated that they read more than half of the text messages, and 63% stated they they tried more than half of the TipsByText activities with their child (see Figure 2.5). Furthermore, 44% of respondents stated that they continued to use several of the activities after the end of the intervention, and 76% of respondents said that they would recommend TipsByText to a friend with children, while only 9% stated that they would not. Thus, the majority of the survey respondents read, used and enjoyed the TipsByText intervention.

Because of the non-representative sample of survey respondents, the most informative data from the parent survey is the free text responses to the survey's last questions, which asked parents about why they either would, would not, or did not know whether they would recommend TipsByText to a friend. To this question, 171 parents responded using their own words, and many of them writing elaborate answers with an average of 25

[2] This low response rate was despite us using multiple strategies that have been shown to improve SMS survey participation (Balabanis et al., 2007; Busara Center for Behavioral Economics, 2019). Specifically, we (a) informed participants about the length of the survey in the first SMS, (b) sent out a reminder via text message to all parents who had not yet responded 5 days after the first request, and (c) offered a reward for participation in the form of a lottery for two gift cards to an online book store (value DKK 500/USD 83).

Figure 2.5: Parent survey responses



Notes: The figure shows the proportion of parents who gave a given response to the parent survey. N indicates the total number of respondents who responded to that particular question.

words and a maximum of 180 words. These free text responses gave us great insight into what the parents appreciated or did not appreciate about the intervention. These insightful free text responses also indicate that the positive responses to the survey questions about usage are not driven solely by social desirability bias because many of the responses include information in the free text responses that would only be available to a parent who engaged with the intervention's text messages.

Overall, the free responses from the parents indicate that most respondents enjoyed the

intervention and felt that they learned something from it. For example, of the 133 parents who would recommend TipsByText, 27 mentioned that the intervention provided them with valuable inspiration, 42 mentioned that they though the content of the texts was good, 24 mentioned that the texts acted as a good reminder for them to talk with their children, and 29 parents mentioned that the texts provided them with new knowledge (see Table 2.1 for an overview of the parents' positive responses).[3]

Table 2.1: Parent survey responses to the free text question

Parents mentioned that the TipsByText intervention was…	**Would recommend TbT (N=141)**	**Would not recommend TbT (N=16)**	**Unsure (N=29)**
Simple/easy/specific/not time consuming	39	0	0
Inspiration/new ideas/good alternatives/input/motivation	35	1	0
Content is good/good tips/ good advice/good activities	60	1	2
Reminder/increased awareness/increased focus	24	0	0
New knowledge/informative/ instructive/"aha" experiences	31	1	0

Notes: The table indicates the number of parents who mentioned a given category in their free text responses. 174 parents gave at least one free response. 123 of these parents would recommend TbT, 15 would not recommend TbT, and 28 didn't know whether they would recommend TbT.

To provide a sense of the sentiment of the free text responses, we include some of the free text responses from parents:

> "What TipsByText can do is to make language acquisition a part of everyday life. I work with this as a teacher but I forget to continue when I come home

[3] All free text answers were coded for content by three independent coders. For cases where there was discordance, the first author made the final decision. Responses could include different content and thus receive multiple codes.

to my own children. But the daily text message from TipsByText has ensured that I am reminded of it and then I did some of it that day. It has been really good."

"It is a good and simple tool to help the child learn language. Especially as a first time mother there are many things you don't really think about that you can do of simple things to promote language. Even if you don't use the text immediately you think about what it says. We have often talked about the text messages at home and both dad and mom have used it."

"Short, specific and precise texts with examples. No raised finger [lecturing] but only specific ideas that can be tried immediately without preparation."

In the same sentiment, as answers to the survey's final question about whether the parents had any additional comments about the intervention, many parents wrote things like "Thanks for now. You have been very important," "Good initiative," and "Thank you for your efforts." One parent even wrote nothing except, "I love you." Several parents also asked whether they could continue to receive the texts, or whether they could receive the same texts again. There is thus an overwhelming positive sentiment about the intervention in the free text responses, and it therefore seems unlikely that low compliance with the intervention is the sole explanation of the lack of effect on the children's language development. This conclusion is also confirmed in the data from the qualitative evaluation.

Qualitative evaluation

To learn about how parents responded to TipsByText, and to gain insights into why they responded the way they did, a qualitative evaluation of the intervention was conducted

concurrently with the RCT. This qualitative evaluation was conducted by an independent team led by Zazie Tolmer, Director of CoIntent ApS, and the conclusions of this separate evaluation was written up in a report, "Qualitative evaluation report of TipsByText in Denmark," before the independent team knew the results of the impact evaluation (Tolmer, 2021).[4] The qualitative evaluation is based on in-depth interviews with 40 parents who received the intervention. These respondents were selected based on pre-specified non-random selection criteria from a sample of 119 families/130 children who were randomly drawn from the full RCT sample and excluded from the RCT (see Figure 1.3). The qualitative evaluation also included in-depth interviews with 5 kindergarten leaders, 10 kindergarten teachers, and brief phone interviews with 20 parents who opted out of the intervention.

Overall, the qualitative evaluation concludes that among the 40 parents interviewed, 23 parents could be grouped under "high engagement," 8 could be grouped under "medium engagement," 6 could be grouped under "low engagement," and 3 could be grouped under "limited engagement" (see engagement rubric in Table 2.2).

Because the 40 respondents in the qualitative evaluation were selected (non-randomly) from a randomized list of 119 families, these results allow us to set a lower bound on high parental engagement of at least one parent in 19.3% of the families in the overall sample (23/119) and medium-to-high engagement in 26% of families. These are very lower bounds, as they assume that all of the parents in the randomized qualitative sample who were not interviewed had no engagement with the intervention.

Thus, as concluded with the parent survey data, low engagement with the intervention

[4] All conclusions and quotes included in this section are based on or taken from this report.

Table 2.2: Parent engagement rubric

	READING THE TEXTS	EXPECTATIONS OF THE TEXTS	USE OF THE TEXTS
HIGH	Read the texts regularly (when they arrive to every couple of weeks). Some re-read the texts several times.	Expect to learn something from the texts. This may range from the texts providing one new fun idea to every text providing new and relevant activities to support their child's language development.	The use varies some parents end up using one tip because it offered something new and others use many of the tips. These parents incorporated one to many of the tips into their routine.
MEDIUM	Read the texts, some parents read them regularly others read them when they can.	Expect that they will know the texts content and that it aligns to what they are already doing.	Use the tips to confirm that their child is on track and that they as parents are doing the right things to support their child's language development and/or as a reminder to spend some time supporting their child's language development.
LOW	Skim read the texts or only read a few texts from time to time	Expect the text content to be obvious 'nothing new' and below their child's language level.	Do not reflect on or use the tips.
LIMITED	Limited Ignore the texts and/or delete the texts	Do not think that the texts are for them.	N/A

Notes: The figure is reprinted with permission from "Qualitative evaluation report of TipsByText in Denmark" by Zazie Tolmer.

seems an unlikely explanation of the lack of effect on children's language development. It is, of course, likely that despite these strong qualitative results, many parents did not engage with the intervention at all. Such lack of engagement would bias our treatment effect estimate toward zero. However, in case there was an effect on language skills in the 26% of families in which at least one parent engaged with the intervention, we would expect to see this in the point estimate of the treatment effect.

2.1.4 Did the intervention affect parents' behavior toward engaging in more and more effective language stimulating activities with their children?

Above we have provided evidence that the three necessary conditions for intervention to affect parent's behavior have been met: parents received the intervention, they did not opt-out of the intervention, and at least 26% of parents read and used the intervention's text messages. The question remains of whether the intervention actually affected the parent's behavior, and whether it did so in ways that we expected would be beneficial to the children's language skills.

The qualitative evaluation can also help us answer this question because it asked *how* the parents used the texts. The overall conclusion from this part of the qualitative evaluation is that parents primarily used the texts in three language-related ways:

- Early literacy: letters, books, reading direction, and visiting the library
- Vocabulary: more nuanced and expanded vocabulary, better enunciation
- Language use: focus on dialogue techniques

These three parent-behaviors are closely related to the intention of the intervention, which is laid out in Chapter 1. All these three domains of child-directed speech have been shown to be positively related to children's early language skills.

Below, we have added short excerpts from the qualitative report for each of these domains of parent's use and added quotes to exemplify the data on which the conclusions have been made.

Early literacy

The qualitative report concludes that the TipsBytext intervention "led them [parents] to dedicate more time to reading with their children and reading in more interactive ways by stopping to ask their children questions tapping into their imagination and creating discussions around the reading material." For example, one parent said:

> "so if we read a book, we ask him, what have we read about? What do you think he's going to do? We're trying to build a dialogue with him about what we've read... before the text messages we just read the book, but after that, we started talking to Nikolai about what we have read about." - Maksim, recipient of the TipsByText intervention.

Vocabulary

The qualitative report concludes that the TipsBytext intervention prompted parents to "reflect on their own use of language and how they could further help their children to expand their vocabulary" For example, one parent said:

> "it [TipsByText] is really good because it's stuff you don't normally think about. For example, why just call a flower a flower, instead of calling it a sunflower if it's a sunflower. Or a rose, if it's a rose, to give him a larger vocabulary." - Kristian, recipient of the TipsByText intervention.

Language use

The qualitative report concludes that the TipsBytext intervention led parents to converse more and in different and more diverse ways with their children. The report concludes that "The tips felt achievable, gave them [parents] the confidence to challenge their children, and made them feel more at ease in doing language activities with their children." For example, one parent said:

> "What I do differently is that I am more at ease but may find it easier to talk about many more different things. And develop a dialogue with Otto. Because sometimes when kids talk, you just say 'Oh yes' (laughs). But I actually think we've all gotten better at developing dialogue and asking back and asking for details and developing on the words." - Karen, recipient of the TipsByText intervention.

Overall, the parental language stimulating behavior elevated in the qualitative report (early literacy, vocabulary, and language use) is consistent with behavior that the literature has identified as being effective in supporting children's language development. We can thus conclude that at least in some families, parents report that the TipsByText intervention changed their behavior and did so in desired ways.

Some parents would have done the suggested activities anyway

The free-text responses in the parent survey also show how not all parents changed their behavior after having received the intervention. Specifically, in the free-text responses, 42 parents mention that they already did many of the tips, or that they used the texts mostly

as confirmations that they were already doing the right thing (see Appendix Table A.13 for an overview of the additional categories mentioned in the free text responses). For example, one parent writes:

> "I have read the texts but as I work in a kindergarten none of the received was new knowledge. I would have spoken in the same way with my children regardless, so for us, it hasn't changed anything."

One parent even writes: "I think the survey is missing the question: Would you have done many of the exercises without having received the suggestions?" Thus, one explanation for the lack of positive effects of the TipsByText intervention is that many parents would have done the activities suggested in the TipsByText text messages regardless of whether they had received the text messages.

We believe that a high baseline engagement in language stimulation is likely to have biased the results toward zero. However, in the parent survey's free-text responses, 32 parents also mentioned explicitly that the TipsByText intervention provided them with new knowledge or taught them something new (see Table 2.5). For example, one parent responds:

> "Gaining verbal language is an incredibly important process for children and parents have the biggest task wrt [with respect to, translated from Danish abbreviation "ifht"] paving the way for the linguistic development. Tipsbytext sends useful and briefly formulated information that is straightforward to do. In addition, the parents gain a series of aha experiences wrt what is important when children learn language"

In addition, the qualitative evaluation contains many observations of parents referring to a specific TipsByText tip when arguing for a certain behavior. For example, one parent says:

> "Well, for example, I got one [TipsByText text message] the other day that said, 'When you read with your child, take his finger and read.' So they get an idea of what direction you're reading in. In fact, I never thought about that being important."

Thus, we find it likely that there is a high counterfactual engagement with language stimulating activities in the sample, and that a high engagement with language stimulating activities in the control group helped drive the null result. However, given how the qualitative evidence and survey responses indicate that many parents engaged with the intervention, we find it unlikely that lack of parental behavior change can fully explain the lack of positive effects of the TipsByText intervention on children's language skills.

2.2 Exploring the link between parent behavior and children's language skills

In the sections above, we have shown evidence that parents received the intervention's text messages, that some parents read and used the text messages, and that some parents used the intervention in ways that the literature has identified as being effective in supporting language growth. Going back to the theory of change presented in Figure 2.1, we have thus confirmed that the first step – the step from the TipsByText intervention to parent's behavior change – was satisfied for some of the parents. We are, therefore, left with the question of whether, and in that case why, parents' changed behavior did not improve children's language skills.

In this section, we will first explore whether there was a link between exposure to TipsByText and growth in children's language skills in the highly selected group of families where parents explicitly expressed engagement with the TipsByText intervention. As we show that this does not seem to be the case, we next explore several options for *why* parents' changed behavior did not improve children's language skills.

2.2.1 Did the intervention improve language skills among children of engaged parents?

One natural question that arises when we find that some parents did engage with, appreciated, and used the intervention in desired ways is whether the intervention was successful in improving the language skills in the families where at least one parent engaged with the intervention.

One obvious challenge in answering this question is the lack of a valid counterfactual for the group of engaged families: we do not know whether the parents in the control group would have engaged with the TipsByText intervention had they received it, and the group of engaged parents does not constitute a representative sample of the full sample of parents. For these analyses, we will thus compare children's language skills at endline to *their own* language skills at baseline across groups. We will ask whether children of engaged parents experienced greater language *growth* over the course of the intervention than did other relevant groups of children (these groups will be defined below).

Overall, we have an indication that 241 parents engaged with the intervention. This data comes from three data sources: parents who in the parent survey responded that they read more than half the text messages, parents who the qualitative evaluation categorized as be-

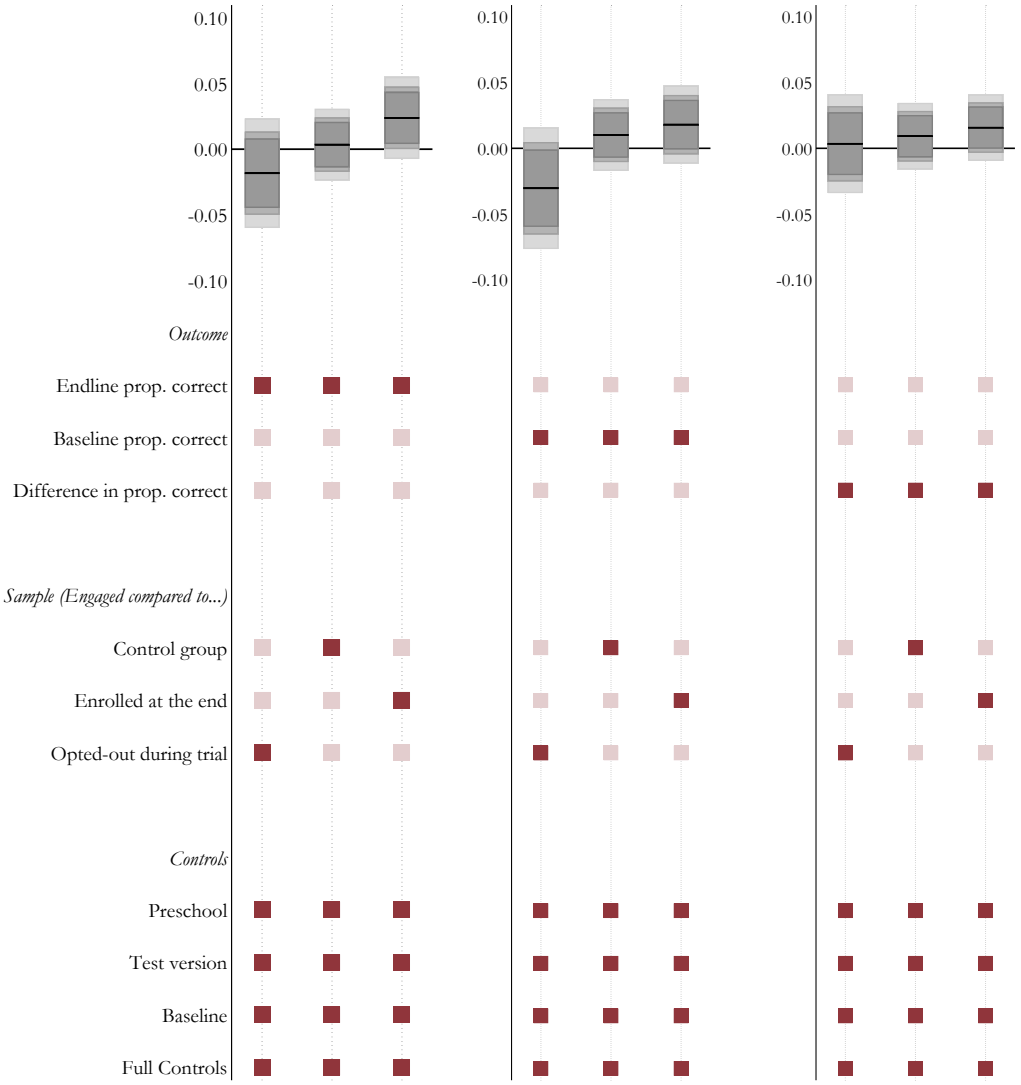
ing highly or moderately engaged in the qualitative evaluation, and parents who at some point responded to the phone number from which the intervention was sent. The language growth of children of engaged parents will be compared to the that of children in three other groups: (1) children of parents who opted out of the intervention, (2) children in the treatment group whose parents did not indicate engagement with the intervention in any way but who did not opt out, and (3) all children in the control group.

As shown in Figure 2.6, the children of the engaged parents do experience a slightly higher language growth than the children in each of the three comparison groups (see Appendix Tables A.17–A.19 for regression results and Appendix Tables A.14–A.16 for more comprehensive balance tables). However, these differences are very small in both relative and absolute terms: across comparisons, there is a maximum difference between groups of 1.3 percentage points in the proportion of correctly answered questions, or 8.7% of a standard deviation of the synthetic control group. Furthermore, none of the difference-in-difference estimates are statistically significant on any conventional significance levels.[5]

These exploratory analyses indicate that the intervention was not successful in improving the children's language skills even in families where at least one parent engaged with the intervention. This is a very surprising result given that the qualitative evaluation concludes that parents who engaged with the intervention tended to use it in ways that the literature has shown to be effective in supporting children's language skills. The next section will thus explore reasons why the intervention might not have been successful in improving children's language skills.

[5] As seen in Appendix Tables A.17–A.19 the point estimates change slightly when controlling for the baseline values, which is consistent with the fact that language skills generally grow faster at lower baseline levels, but the conclusions of very small differences between groups do not change.

Figure 2.6: Differential language growth for children of engaged parents compared to other groups of children



Notes: The figure shows the coefficients and confidence intervals for different regression specifications of the difference in outcomes between the group of children of parents who have indicated engagement with the intervention and other groups indicated by "Sample (Engaged compared to...)" Coefficients are depicted by black horizontal lines. The vertical bars, from lightest to darkest, denote the 90%, 95% and 99% confidence intervals, respectively. The bottom panel indicates the regression specification in terms of the outcome measure, the sample, and the set of control variables.

2.2.2 Why did parental behavior not improve children's language skills?

If we are ready to accept that some parents did in fact use the intervention, that the intervention did in fact change some parents' behavior in ways that the literature has identified as being beneficial to language development, and that the intervention did not improve language skills, as measured by the language assessment, "Sprogvurdering 3–6," even among the children of engaged parents, we must conclude that children's performance in the language assessments simply did not benefit from the parental behavior change.

The TipsByText intervention is designed around the large and robust literature in child development establishing that the quantity and quality of child-directed speech predict children's early language development. In this section, we will return to the literature to explore some possible reasons why child-directed speech does not seem to improve language skills in this study. In particular, we will discuss whether different characteristics of the sample (children's age, their kindergarten hours, or their baseline language skills) or characteristics of the language assessment might have hindered the effect.

As we do not have sufficient data to be able to either reject or fail to reject any of these possible explanations for the null effect, this section should be regarded as an exploratory discussion.

Characteristics of the sample: The children's age, long hours in kindergarten, or strong baseline language skills

The period most important for shaping children's language abilities is argued to be below age 3 (Schwab and Lew-Williams, 2016; Cunha and Heckman, 2007). Thus, most literature linking child-directed speech and children's language development focus on establish-

ing a relationship at this early age (Fernald and Marchman, 2012). The TipsByText trial in Denmark focused on 3–5-year-old children for two reasons. First, the original TipsByText intervention, READY4K! focused on preschool children with an average age of 4.4 years. Second, the children in Denmark typically attend kindergarten from age 3 to age 5, and the collaboration with the municipalities happened at the kindergarten level. Thus, one possible explanation for the lack of effect is that parents' impact on children's language growth is not as strong when the children are older.

This argument could be strengthened by the fact that children in Denmark are among the children in the world who spend most hours per day in kindergarten at an average of seven hours (Villadsen and Hviid, 2017). So perhaps in our sample, the "low-hanging fruits" of language stimulation, which a light-touch intervention like TipsByText would be able to reap, might already have been reaped by the kindergarten.

Related, some parents also mentioned in the parent survey and the qualitative evaluation that the content of the text messages was too easy for their children's level. If the children's language skills are so strong at baseline that the tips in the intervention are not new to them, this could possibly explain the lack of positive effect.

However, a strong empirical argument against these child characteristics (children's age, hours in kindergarten, or strong baseline skills) explaining the lack of effect of the intervention is that there is no differential treatment effect in any of the relevant groups. Particularly, there are no positive treatment effects among the youngest children (Appendix Table A.20) or among the children who scored in any of the two bottom quartiles in the baseline language assessment (Appendix Table A.21).[6] There is also no differential treatment effect

[6] If anything, there actually seems to be a marginally significantly negative effect of the children who scored in the second quartile of the baseline language assessment, and among the 3-year-olds. However, note

for the children who has a parent who is working part-time or not in the labor force (see Appendix Table A.22). Thus, the children's age, their long hours in kindergarten, or strong baseline language skills are not valid contenders to explain the lack of positive effect of the intervention.

Characteristics of the language assessment

Another possible reason for the lack of measured effect of the intervention is that the language assessment instrument used at both baseline and endline, "Sprogvurdering 3–6," is either not sensitive enough to capture changes in language skills, or that it does not measure the dimensions that are affected by the intervention.

As mentioned in Chapter 1, the language assessment instrument measures children's language skills in 4-to-7 dimensions depending on the age of the child (see Table 1.4 for an overview of the dimensions). The instrument is developed by child development researchers Dorte Bleses, Anders Højen, Guido Makransky, Philip Dale, and Werner Vach as a preventive population language screening instrument. The instrument is a further development of the screening instrument, "Screening Instrument for 3-Year-Olds" (SI-3), which has been found to have acceptable psychometric properties for most subscales and to be sufficiently negatively skewed to differentiate among children with low scores (Bleses et al., 2010). Furthermore, researchers from the ROCKWOOL Foundation recently established that the children's scores in the language assessment correlate with their scores in a standardized Danish test in second grade; for example, 37% of the children who score in the bottom quintile of the language assessment also score in the bottom quintile

here that analysis on 3-year-olds is based on only 85 children, because it includes only the few children who started kindergarten before they turned three.

in 2nd grade, whereas only 6% of this group score in the top quintile in second grade (see http://udvikling.rff.dk). Although the goal of this study was not to establish the external validity of the language assessment, this correlation suggests that the assessment does have some external validity in one dimension we care about: Test scores in school.

With regards to whether the assessment is not sensitive enough to capture effects of an intervention, we are aware of at least two large-scale RCTs that have used the instrument as the main outcome measure. Both interventions are primarily kindergarten-based, conducted in a general population, and cluster-randomized on the kindergarten level. One intervention is a storybook-based intervention called "Structured Preschool Efforts in Language and Literacy" (SPELL), and the other is a scripted lessons intervention with a language/literacy focused curriculum called LEAP. Both interventions found significant treatment effects of $0.21 - 0.32$ S.D. on pre-literacy skills (rhyme detection, deletion, letter identification, and sound discrimination) but no significant effects on other dimensions of language skills, such as vocabulary, comprehension, and communication (Bleses et al., 2018a,b). It thus seems that at least the pre-literacy domains of the instrument are sensitive enough to capture effects of an intervention.[7]

With regards to whether the assessment instrument does not measure the language dimensions affected by the TipsByText intervention, the intervention was explicitly designed around the dimensions included in the assessment, as these are the dimensions that the Ministry for Children and Education has emphasized as important areas of focus for the kindergartens (Bleses et al., 2010). If we were to see effects on the children's language skills,

[7] It should be noted here that the two interventions are both kindergarten-based and randomization has happened on a kindergarten level. Given that the assessment is conducted by kindergarten staff the testers cannot have been blind to the treatment, as they were in our study, which might have affected the outcome.

we would therefore expect to see them exactly on the language dimensions measured by the instrument "Sprogvurdering 3–6."

The empirical link between child-directed speech and language skills

The TipsBytext intervention is motivated by the large and robust literature in child development establishing that the quantity and quality of child-directed speech predict children's early language development. However, the TipsByText intervention does not improve children's language skills despite affecting parents' behavior in ways that have been identified in the literature as being effective in supporting language growth. The above discussion about the characteristics of the sample and the language assessment does not provide a conclusive explanation for this lack of positive effect. A possible and provocative interpretation of the lack of positive effect of the TipsByText intervention on children's language skills is that the correlation between child-directed speech and children's language skills is actually not causal – at least not *universally* causal across age groups or baseline language skills.

Most of the literature linking child-directed speech to children's language skills is based on correlations between a measure of child-directed speech and a measure of children's language skills. Child-directed speech is often measured with the advanced electronic tool, LENA, which counts the number of utterances spoken to the child and is able to analyze the type of utterances (see https://www.lena.org/). The correlation between child-directed speech and children's language skills is thus based on data of very high quality. However, evidence that this link is causal is sparse.

Many lab experiments have shown that certain characteristics of child-directed speech

casually affect children's learning of new words. However, few studies have manipulated child-directed speech in the field and studied how this manipulation affects language skills.

The study with the most potential to inform on this causal relationship is the evaluation of the "30 Million Word Gap Project" by Dana Suskind and colleagues (Suskind et al., 2016). This project includes an intervention to the quality and quantity of child-directed speech and finds positive effects on parental behavior. However, the pilot study does not include outcome measures on children's language skills. Thus, the study has shown that it is possible to affect parent's behavior in the field, but not that this behavior change, in turn, improves children's language.

This study alone does not provide sufficient data to make a conclusion about whether the link between child-directed speech and children's language skills is causal. However, the interpretation is consistent with the data and there is not strong evidence from the literature to refute this interpretation. Thus, one important implication of this study is that manipulating the quality and quantity of child-directed speech might not be sufficient to improve children's language skills in a diverse sample of children and their parents.

2.3 Conclusion

In this chapter, we discuss several different reasons why the TipsByText intervention was not successful in improving children's early language skills. We show that the lack of positive effect is not driven by the intervention not having been delivered successfully, as more than 90% of parents received all the TipsByText text messages, and as 83.7% of children still had at least one parent enrolled at the end of the intervention period. We also show evidence that many parents read and used the text messages, and we calculate a very conser-

vative lower bound for medium-to-high engagement of 26% of families. Finally, we present qualitative evidence that at least some parents used the tips in a way that the literature has identified to be beneficial for language stimulation, by dedicating more time to reading with their children, creating discussions about the reading material, increase the complexity of the vocabulary they used with their children, and converse more, and in more diverse ways, with their children. Thus, lack of successful intervention delivery or lack of parental engagement with the intervention does not seem to be able to account for the lack of positive effect of the intervention on children's language skills.

In the chapter, we acknowledge that the effect of the intervention is likely biased toward zero because not all parents used the intervention (even if they did not opt-out), and that some parents already did the activities suggested by TipsByText. The effect might also be biased toward zero because the children in the sample are at a stage in language development where the quantity and quality of words they hear from their parents affect their language growth less, and because they spend many hours per day in kindergarten away from their parents. However, we show suggestive evidence that even among the parents who used the intervention, there seems to have been no effect on the children's language skills, which indicates that the intervention was not successful in improving children's language skills even in the families who used the intervention.

We explore whether the lack of positive effect can have been partially driven by the fact that the language assessment instrument, "Sprogvurdering 3–6" is not sensitive enough to detect small changes in language skills. However, we conclude that this explanation is unlikely to be the sole explanation of the lack of positive effect because the TipsByText intervention is designed to explicitly address the language dimensions measured with the

instrument, because the instrument is designed to be skewed toward the low end of the distribution, and because other studies have found positive effects of interventions using the same instrument.

We are, therefore, left with two possible explanations for why the TipsByText intervention did not improve children's language skills. The first explanation is that the lack of positive effect of the intervention is not caused by a single factor but by a multitude of different biases toward zero, such as many parents not engaging with the intervention, parents not changing their behavior enough to impact their children's language development, the children being too old to experience large benefits of the way their parents speak to them, etc. This possible explanation assumes that the TipsByText intervention could have had a positive impact on children's language skills under different circumstances, such as e.g. an opt-in design which might have increased parental engagement with the intervention, a sample of younger children, or a different outcome measure. We cannot rule out this explanation, but the tightly estimated null effect is evidence against this explanation, as this means that even if the intervention did have a positive effect on some children's language skills, there must have been another group for whom the effect was similarly negative.

The second explanation, which can follow directly from the first, is that children's language skills are not easily changed with a light-touch parent-oriented intervention when it is implemented in a large-scale effectiveness study with a large and diverse sample. This interpretation suggests that leveraging parents to improve children's early life skills might not be as effective as a large-scale policy tool as the literature suggests.

This second explanation is consistent with the conclusion that the strong correlation between the quality and quantity of child-directed speech and children's language skills found

in the literature is not universally causal, meaning that a manipulated increase in the quality and quantity of child-directed speech does necessarily not lead to improved language skills. This possible explanation assumes that the TipsByText intervention could *not* have had a positive impact on children's language skills under different circumstances. The results of this study are consistent with this explanation, but it is important to note that this study was not designed to answer the broader question of causality between child-directed speech and language skills, and thus it should not be interpreted as evidence against causality. Instead, it does open interesting questions for further research: Is there a causal effect of the quality and quantity of child-directed speech on children's language skills in naturalistic settings and for typically developing children outside the lab?

3

Discussion

Early-life skills are a strong predictor of later life outcomes, and parents play an instrumental role in their children's acquisition of early-life skills. However, there is still a question in the literature about the effectiveness of leveraging parents to improve children's early-life skills as a large-scale policy tool.

In this dissertation, we estimate the impact of a parent-oriented intervention, TipsByText, in a naturalistic setting with a large and diverse sample. TipsByText aims at increasing language skills among preschool children by sending text messages directly to their parents with tips about how to incorporate simple language stimulation exercises into their everyday lives. The intervention is both low-cost, easy to scale, and broad in scope, by covering many domains of language development. The TipsByText curriculum is based on evidence from child development research on effective ways to support children's language development.

To evaluate the effectiveness of the TipsByText intervention, we use data from a randomized controlled trial (RCT) conducted in Denmark in 2019/2020 among 3,305 families with 3,804 three-to-five-year-old children. Parents in the treatment group received the TipsByText intervention for eight months, while parents in the control group received no intervention.

In Chapter 1, we show that the TipsByText intervention was not successful in improving language skills in our sample. We show that the intervention did not affect average language skills in the treatment group, that it did not affect any of the dimensions of language skills measured by "Sprogvurdering 3-6," and that it did not improve the language skills of any of the groups of children whom we, based on the current literature, had expected would benefit the most from the intervention.

In Chapter 2, we discuss potential explanations for why the TipsByText intervention did not improve language skills. We show that (1) the lack of effect is unlikely to be a result of the parents not having received the intervention's text messages because more than 90% of parents received more than 90% of the text messages and because there does not seem to be

differences in children's language skills depending on how many text messages the parents received, and (2) the lack of effect is unlikely to be a result of parents not having used the intervention because qualitative interviews reveal that many parents used and appreciated the intervention and because there seem to be no differences in language growth between children of parents who expressed engagement with the intervention and children of parents who did not. Therefore, we conclude that there are two possible explanations for why the TipsByText intervention failed to improve children's language skills. The first explanation is that the lack of positive effect of the intervention is caused by a multitude of different biases toward zero, such as many parents not engaging with the intervention, parents not changing their behavior enough to impact their children's language development, the children being too old to experience large benefits of the way their parents speak to them, etc. We argue that the tightly estimated null result provides empirical evidence against the first explanation, but we cannot fully rule it out. The second explanation, which can follow directly from the first, is that children's language skills are not easily changed with a light-touch parent-oriented intervention when it is implemented in a large-scale effectiveness study with a large and diverse sample. This interpretation suggests that leveraging parents to improve children's early-life skills might not be as effective as a large-scale policy tool as the literature suggests.

3.1 Limitations of the current study

This dissertation provides an example of an evidence-based parent-oriented intervention that was not successful in improving children's language skills in a large and diverse sample. However, the study is not without its limitations.

The most important limitation of the current study is that we do not have representative data on parent's behavior either before or after the intervention. We have retrospective survey data on parents' use of the intervention from one month after the end of the intervention, but only 9.4% of the children in the treatment group had at least one parent who responded to this survey. We also have qualitative data on parental use of the intervention from the concurrent qualitative evaluation. However, even though this data is valuable in informing us about how parents used the intervention, it is based on a small and non-representative sample, and we do not have similar data from the control group. Ideally, we would have been able to collect data on parent-child interactions at endline in both the treatment and control group, using either the MacAuthur-Bates Communicative Development Inventories (MCDI, see https://mb-cdi.stanford.edu/), which is a language inventory filled out by parents, or the Language ENvironment Analysis tool (LENA, see www.LENA.org), which is a technological device capable of counting and analyzing words and utterances in parent-child interactions. Such data would allow us to make the relevant empirical distinction between whether the lack of positive effect of the TipsBytext intervention is driven by the intervention not causally affecting how and how much parents talked to their children, or whether it is driven by the fact that the changes in child-directed speech did not causally affect children's language skills. The answer to this question would be of both theoretical and empirical relevance, but unfortunately, collection of representative parent behavior was beyond the scope of this study.

A second limitation of the current study is that the language assessment tool used to assess children's language skills at baseline and endline is only used in Denmark. Although an earlier version of this tool has been found to have acceptable psychometric properties for

most subscales (Bleses et al., 2010), the fact that the tool is only used in Denmark makes it impossible to compare the results and effect sizes of this study with other studies that use alternative language and more internationally validated assessment tools. Furthermore, the use of a non-standard assessment tool limits our comparisons of baseline language skills between our sample and other samples. Thus, we cannot analyze whether the lack of positive effects of the intervention could be driven by, for example, the current sample having higher baseline language skills than in other studies which find positive effects of parent-oriented language intervention.

A third limitation of the current study comes from its timing: The TipsByText intervention lasted eight months from late January 2020 to early September 2020, which means that part of the intervention coincided with the spread of the COVID-19 virus and the lockdown associated with the first wave of the pandemic. The pandemic did not affect the implementation of the RCT, and the randomization ensures that in expectation, the lockdown affected the treatment and control groups the same way. Thus, the internal validity of the results should be intact despite the pandemic. However, the pandemic and the associated lockdown might have challenged the external validity of the results presented in this dissertation. Below, we will argue that the direction of any external bias of the effect of TipsByText is ambiguous and that we are more concerned with the precision of the estimate than with its bias. Thus, given that the estimated impact of the TipsByText intervention is a tightly estimated null effect, we are not concerned that the lack of effect can be explained by the timing of the RCT, and we trust the external validity of the null result to replicate to non-COVID times.

3.1.1 Denmark's response to the Covid-19 crisis

The Danish government was among the first in Europe to declare a national lockdown (Olagnier and Mogensen, 2020). Specifically, from March 16th to April 15th, all schools and preschools were closed, non-essential public employees were sent home, and all privately employed were encouraged to stay home if possible.[1] For this reason (and others), Denmark was hit relatively mildly by the first wave of the virus: with a population of 5.8 million people, Denmark has experienced 723 deaths of citizens diagnosed with COVID-19 as of November 3rd, 2020, whereas one of Denmark's closest neighbors, Sweden, with a population of 10.3 million people, has experienced 5.923 deaths.[2]

In summary, Denmark enacted an "Act fast and act with force" strategy that resulted in the country being severely affected by the lockdown for a short period but less affected by the pandemic itself compared to other, similar countries.

3.1.2 How the pandemic and associated lockdown possibly affected the Tips-ByText evaluation

The randomization ensures that the average circumstances of families in the treatment and control groups were affected equally by the pandemic. However, children's language skills might not be affected equally in the treatment and control group if parent's circumstances interacted with their treatment assignment to determine their level of engagement in active

[1] On March 11, the government announced a minimum two-week lockdown: https://nyheder.tv2.dk/samfund/2020-03-11-danmark-lukker-ned-her-er-regeringens-nye-tiltag, retrieved on November 3, 2020. On March 23, the government extended the lockdown by another two weeks: https://www.dr.dk/nyheder/politik/mette-frederiksen-danmark-vil-fortsat-vaere-lukket-ned-frem-til-13-april, retrieved on November 3rd, 2020.

[2] Numbers extracted on November 3, 2020, from https://www.statista.com/statistics/1093256/novel-coronavirus-2019ncov-deaths-worldwide-by-country/

language stimulation with their children.

We believe the lockdown can have changed parents' engagement in active language stimulation in two opposite directions. First, due to the closure of preschools and physical workplaces, we expect that parents have spent more time with their children during the lockdown. Such a parent-child time-shock could increase parental engagement in active language stimulation for both the treatment and control group above the pre-lockdown level. At the same time, during the lockdown, parents had to juggle homeschooling, working from home, social restrictions, and for some parents, the health effects of the pandemic itself. Thus, despite the possible positive time-shock, parents might have experienced a negative shock to cognitive bandwidth, which could decrease the parental engagement in active language stimulation for both the treatment and control group below the pre-pandemic level.

We find it plausible that the pandemic affected the time and bandwidth of parents in both the treatment and control groups. However, the pandemic affecting parental behavior is not inherently challenging for the external validity of the effect size. In fact, if children's language skills are affected only through a *parallel shift* in the distribution of the outcome measure, y_i, the intersect from Equation 1.1, $\hat{\alpha}$, will fully capture the parallel shift among the control group, and the estimate of the treatment effect, $\hat{\beta}$, will still be an unbiased estimate of the non-pandemic effect size of the intervention.

Whether the change in parent behaviors created an external bias in the estimated effect size depends on whether the pandemic *differentially* affected the language skills of children in the treatment and control groups. Here the answer is ambiguous. Parents in the treatment and control groups are expected to have experienced the same time shock. However,

for a given time investment in language stimulation, we would expect the *quality* of the increased investment to be higher for parents in the treatment group, which would bias the effect upward compared to a non-pandemic counterfactual. On the other hand, a decrease in parental bandwidth might have decreased the number of text messages the parents in the treatment group read, which would bias the effect toward zero.

Thus, we argue that it is unlikely that the estimate of the effect of the intervention is systematically biased in one direction, and it is unclear whether the opposing challenges result in an *average* upward or downward bias of the effect compared to a situation with no lockdown and no global pandemic. What we think is more likely is that the pandemic and the associated lockdown increased the *dispersion* of the endline language scores, as we expect the pandemic to have affected different families differently independently of their treatment status. For example, the total effect of the lockdown on parental engagement in language stimulating activities might be positive for a family where one parent is a stay-at-home parent, whereas the total effect might be negative for a family where the parents had to deal with hospitalized grandparents. To see how differential effects of the pandemic can increase dispersion in endline language scores, remember that the standard errors of the effect size estimated from Equation 1.1 are proportional to the residual, $\hat{\varepsilon}_i = y_i - \hat{y}_i$, which means that a wider dispersion on endline language scores translates into larger standard errors in the main analysis. Larger standard errors, in turn, will bias us toward a type 1 error, in which we reject a true hypothesis.

To add to the uncertainty about the direction of the bias, we do not know whether parental and preschool language stimulation interact in the children's language production function. If parental and preschool language stimulation are substitutes, the closure

of preschools should affect children in the treatment and control groups equally. However, if parental and preschool language stimulation complement each other, children in the treatment group might experience a relatively lower benefit of the intervention during the lockdown, which would attenuate the estimated effect of the intervention compared to a situation with no lockdown.

Thus, the COVID-19 pandemic and associated lockdown potentially increased the dispersion of the outcome measure, as families were affected differently by the lockdown. However, given that the estimated effect size of the TipsByText intervention is a tightly estimated null effect, we believe that the lockdown did not significantly affect our conclusions or the external validity of this study.

3.2 Possible positive impact beyond improving children's language skills

So far in this dissertation, the only outcome measure we have focused on is children's language skills; the intervention was designed explicitly to improve language skills and hence our hypotheses focused on language skills. However, our communication with the parents in the treatment group has suggested that the intervention might have positively affected other outcomes than language skills. Below, we will list some of these potential benefits. We do not have data to support or refute whether the intervention causally affected these outcomes, so this section should be seen as purely speculative and could be used for inspiration for outcomes to include in future studies of parent-oriented interventions.

Parents' cognitive bandwidth

Data from the parent survey and the qualitative evaluation reveal that several parents in the treatment group felt that the TipsByText intervention provided them with inspiration for language activities to do with their children (see Table 2.5 and Tolmer, 2021). Thus, it is possible that even though the intervention did not improve children's language skills, it provided valuable support for the parents, which, in turn, might have freed up some of their cognitive bandwidth – a scarce resource for parents of preschool children. For example, one interviewed parent said:

> "[...] you can read and borrow and buy a ton of books about children's development – but there isn't always the time to read them all, - it [TipsByText] is simple and short, and some things you can translate into action!" – Jeanette, recipient of the TipsByText intervention.

Another parent mentioned that TipsByText gave her an opportunity to slow down in a hectic day:

> "[...] the texts created a break, an opportunity to slow down in a hectic day. I think it helped me [...] often my days are very structured and very planned, but these text messages broke up the day for me, and then you can think a little out of the box, I think that's been great." – Ane, recipient of the TipsByText intervention.

The same parent later adds:

> "I was feeling really down for quite a while, and during that time I used them [the text messages] a lot. When you are feeling down you have less energy and time for your kids. And then you feel bad about that too and it just becomes a vicious cycle [...] these text messages have been part of helping me to feel good again."

These parent sentiments indicate that even though the intervention turned out to not benefit children's language development, it might have provided a benefit for the parents. This possibility is best captured in the parent quote that opened Chapter 2:

> "It's a good question whether it has made a difference for my child. I don't necessarily think so. But it's made a difference to ME." – Birgitte, recipient of the TipByText intervention.

Quality of parent-child interactions, social skills, and family routines

Data from the parent survey and qualitative evaluation also reveal that several parents found TipsByText's activities "fun" (see Table 2.5, and 12 out of 31 parents in the qualitative evaluation specifically spoke of the tips being fun). If the intervention can help parents see language stimulating activities as "fun," it is possible that the intervention can have contributed to positive parent-child interactions even if these interactions did not benefit the child's language skills. For example, one parent said:

> "We teach her Arabic letters, but the normal way is like 'this letter blah blah blah.' But a good thing in TipsByText is making it like a game. This was very useful for us." – Mohammed, recipient of the TipsByText intervention.

Another parent tells a story about how the family made a game out of building on a word together: The child said the word "apple," the father responded "pomegranate" (which in Danish is a compound word consisting of "grenade" and "apple"). The son said "grenade" and the father responded "hand grenade," and they all laughed (Tolmer, 2021).

Parents also spoke of how the TipsByText intervention has inspired them to "focus more on good routines and building their children's broader social skills" (Tolmer, 2021). For example, one parent mentioned how a tip about talking to the child while they are getting dressed had changed her view on whether the child should get dressed himself:

> "In the kindergarten, they have told me many times 'He should put on clothes himself.' And I said 'No, he is so small! Why should he do it on his own, I can help him, or my husband can help him.' But when I got the text, I changed my mind. After that, I thought, 'OK, they are right."' – Morsal, recipient of the TipsByText intervention.

Several parents also mentioned that the TipsByText activities were a good alternative to allowing the child to play with an iPad or other electronic devices.

Thus, it seems that some parents have experienced benefits of the TipsByText intervention independently of their children's language development – benefits that might have a value independently of the effects on children's language skills.

The parental engagement with and potential benefit of the intervention opens up an interesting policy question: Should an intervention that is as cheap to implement and easy to scale as the TipsByText intervention be discarded fully for not having achieved what it was designed to do, which was to improve children's language skills, even if it turns out to have unintended benefits? It is the opinion of the first author that the present impact evaluation

does not support the TipsByText intervention being rolled out in Denmark as an evidence-based language stimulation intervention. However, it could still be offered to parents of preschool children, if it were framed instead as an intervention to provide parents with inspiration for fun and easy-to-implement activities to do with their child around language. However, with the current data, it is impossible to say whether parents would be interested in taking up and using the intervention if they knew that the intervention likely would not improve their children's language skills.

3.3 Future directions

The present impact evaluation shows that the TipsByText intervention did not improve children's language skills, as measured by the language assessment "Sprogvurdering 3–6," within one year of the intervention's start date. However, despite this lack of positive short-term effects, it is possible that the intervention has sowed the seeds for a different language culture in the treatment families; a language culture that could create longer-term benefits. For example, if the TipsByText intervention has encouraged some families to have more fun with language at home, as some of the above quotes suggest, the children might have internalized that learning language is fun, which might, in turn, benefit the child in later learning contexts. Indeed, a recent study of the long-term benefits of attending preschool has found significant long-term benefits of attending preschool in terms of e.g. high school graduation and college enrollment, while preschool enrollment did not improve the children's performance on standardized tests (**?**). There is, therefore, a possibility that the TipsByText intervention can have longer-term positive effects on language skills despite the lack of short-term effects. For this reason, we plan to follow the children in the RCT into the

future, which is made possible by the link of the RCT data to the Danish register data. The first follow-up study will be conducted in the Spring of 2022 after the language skills of the children in the RCT who have not yet entered first grade will once again be assessed with the language assessment "Sprogvurdering 3–6." In later years, we will be able to observe the children's grades in school; currently, all children enrolled in public primary schools are tested in Danish in second, fourth, sixth, and eighth grade. Furthermore, related to the question of whether the intervention changed other aspects of the children's lives, such as social skills, we will be able to follow children's well-being in school, which is currently measured yearly in all public schools in Denmark.

References

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey Wooldridge. "When should you adjust standard errors for clustering?" Technical report, National Bureau of Economic Research (2017).

Anderson, Michael L. "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(2008), 1481–1495.

Balabanis, George, Vincent-Wayne Mitchell, and Sarah Heinonen-Mavrovouniotis. "SMS-based surveys: Strategies to improve participation." *International Journal of Advertising*, 26(2007), 369–385.

Bleses, Dorthe, Anders Højen, Philip S. Dale, Laura M. Justice, Line Dybdal, Shayne Piasta, Justin Markussen-Brown, Laila Kjærbæk, and E. F. Haghish. "Effective language and literacy instruction: Evaluating the importance of scripting and group size components." *Early Childhood Research Quarterly*, 42(2018a), 256–269.

Bleses, Dorthe, Anders Højen, Laura M. Justice, Philip S. Dale, Line Dybdal, Shayne B. Piasta, Justin Markussen-Brown, Marit Clausen, and E. F. Haghish. "The effectiveness of a large-scale language and preliteracy intervention: The SPELL randomized controlled trial in Denmark." *Child development*, 89(2018b), e342–e363.

Bleses, Dorthe, Werner Vach, Rune N. Jørgensen, and Torben Worm. "The internal validity and acceptability of the Danish SI-3: A language-screening instrument for 3-year-olds." *Journal of Speech, Language, and Hearing Research*, (2010).

Brooks-Gunn, Jeanne and Lisa B. Markman. "The Contribution of Parenting to Ethnic and Racial Gaps in School Readiness." *The Future of Children*, 15(2005), 139–168.

Bruhn, Miriam and David McKenzie. "In pursuit of balance: Randomization in practice in development field experiments." *American economic journal: applied economics*, 1(2009), 200–232.

Busara Center for Behavioral Economics. "Getting the most out of your SMS Survey: Experimental Results." (2019).

Børne- og Undervisningsministeriet. "Vejledning til Sprogvurdering 3-6." Technical report (2019).

Canfield, Caitlin F., Anne Seery, Adriana Weisleder, Catherine Workman, Carolyn Brockmeyer Cates, Erin Roby, Rachel Payne, Shari Levine, Leora Mogilner, Benard Dreyer, and Alan Mendelsohn. "Encouraging parent–child book sharing: Potential additive benefits of literacy promotion in health care and the community." *Early Childhood Research Quarterly*, 50(2020), 221–229.

Cunha, Flavio and James Heckman. "The technology of skill formation." *American Economic Review*, 97(2007), 31–47.

Cunha, Flavio, James J. Heckman, Lance Lochner, and Dimitriy V. Masterov. "Interpreting the evidence on life cycle skill formation." In E. Hanushek and F. Welch, editors, "Handbook of the economics of education," volume 1. Elsevier (2006). 697–812.

Davidai, Shai, Thomas Gilovich, and Lee D. Ross. "The meaning of default options for potential organ donors." *Proceedings of the National Academy of Sciences*, 109(2012), 15201–15205.

Dweck, Carol S. *Mindset: The new psychology of success*. Random House Digital, Inc. (2008).

Farkas, George and Kurt Beron. "The detailed age trajectory of oral vocabulary knowledge: differences by class and race." *Social Science Research*, 33(2004), 464–497.

Fernald, Anne and Virginia A. Marchman. "Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers." *Child development*, 83(2012), 203–222.

Fernald, Anne, Virginia A. Marchman, and Adriana Weisleder. "SES differences in language processing skill and vocabulary are evident at 18 months." *Developmental science*, 16(2013), 234.

Gelman, Andrew and Jennifer Hill. "Sample size and power calculations." In "Data analysis using regression and multilevel/hierarchical models," Cambridge University Press (2007).

Gilkerson, Jill, Jeffrey A. Richards, Steven F. Warren, D. Kimbrough Oller, Rosemary Russo, and Betty Vohr. "Language experience in the second year of life and language outcomes in late childhood." *Pediatrics*, 142(2018).

Golinkoff, Roberta Michnick, Kathy Hirsh-Pasek, and Dorothy G. Singer. *Why play = learning: A challenge for parents and educators*. Oxford University Press (2006).

Hart, Betty and Todd R. Risley. *Meaningful differences in the everyday experience of young American children.* Paul H Brookes Publishing (1995).

Heckman, James J. "The case for investing in disadvantaged young children." *CESifo DICE Report*, 6(2008), 3–8.

Heidlage, Jodi K., Jennifer E. Cunningham, Ann P. Kaiser, Carol M. Trivette, Erin E. Barton, Jennifer R. Frey, and Megan Y. Roberts. "The effects of parent-implemented language interventions on child linguistic outcomes: A meta-analysis." *Early Childhood Research Quarterly*, 50(2020), 6–23.

Hirsh-Pasek, Kathy, Lauren B. Adamson, Roger Bakeman, Margaret Tresch Owen, Roberta Michnick Golinkoff, Amy Pace, Paula K. S. Yust, and Katharine Suma. "The contribution of early communication quality to low-income children's language success." *Psychological Science*, 26(2015), 1071–1083.

Hoff, Erika. "The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech." *Child development*, 74(2003), 1368–1378.

———. "Interpreting the early language trajectories of children from low-SES and language minority homes: Implications for closing achievement gaps." *Developmental Psychology*, 49(2013), 4–14.

Hoff-Ginsberg, Erika. "Mother-child conversation in different social classes and communicative settings." *Child Development*, 62(1991), 782–796.

Huttenlocher, Janellen, Wendy Haight, Anthony Bryk, Michael Seltzer, and Thomas Lyons. "Early vocabulary growth: relation to language input and gender." *Developmental psychology*, 27(1991), 236.

Kragh-Müller, Grethe. "The key characteristics of Danish/Nordic child care culture." In "Nordic social pedagogical approach to early years," Springer (2017). 3–23.

Kragh-Müller, Grethe and Rebecca Isbell. "Children's perspectives on their everyday lives in child care in two cultures: Denmark and the United States." *Early Childhood Education Journal*, 39(2011), 17–27.

Leffel, Kristin and Dana Suskind. "Parent-directed approaches to enrich the early language environments of children living in poverty." In "Seminars in speech and language," volume 34. Thieme Medical Publishers, 267–278.

Ludwig, Jens, Sendhil Mullainathan, and Jann Spiess. "Machine-learning tests for effects on multiple outcomes." *arXiv preprint arXiv:1707.01473*, (2017).

———. "Augmenting pre-analysis plans with machine learning." In "AEA Papers and Proceedings," volume 109. 71–76.

National Early Literacy Panel. "Developing Early Literacy: Report of the National Early Literacy Panel." *Washington, DC: National Institute for Literacy*, (2009), 260.

Nielsen, Hanne, Peter Jensen, Dorthe Bleses, Jeppe Ostersen, Line Dybdal, and Sanni Breining. "Forskningsbaseret evaluerings-rapport om Fremtidens Dagtilbud undersøgelse af indsatsens implementering og effekter." *København: Børne-og Socialministeriet*, (2017).

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

Olagnier, David and Trine H. Mogensen. "The Covid-19 pandemic in Denmark: Big lessons from a small country." *Cytokine & Growth Factor Reviews*, (2020).

Pepper, Jan, Elaine Weitzman, and Hanen Centre. *It takes two to talk: A practical guide for parents of children with language delays*. The Hanen Centre (2004). Google-Books-ID: Iyj23q3LoiUC.

Roberts, Megan Y. and Ann P. Kaiser. "The effectiveness of parent-implemented language interventions: A meta-analysis." (2011).

ROCKWOOL Foundation. "RFI Indsigt: TipsByText - SMS'er for at fremme børns tidlige sprogudvikling." Technical report (2019).

———. "RFI Indsigt: TipsByText - Tilpasning af en sprogindsats fra USA til Danmark." Technical report (2021).

Rowe, Meredith L. "Understanding socioeconomic differences in parents' speech to children." *Child Development Perspectives*, 12(2018), 122–127.

Rowe, Meredith Lee. "Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill." *Journal of child language*, (2008).

Schwab, Jessica F. and Casey Lew-Williams. "Language learning, socioeconomic status, and child-directed speech." *Wiley Interdisciplinary Reviews: Cognitive Science*, 7(2016), 264–275.

Suskind, Dana L., Kristin R. Leffel, Eileen Graf, Marc W. Hernandez, Elizabeth A. Gunderson, Shannon G. Sapolich, Elizabeth Suskind, Lindsey Leininger, Susan Goldin-Meadow, and Susan C. Levine. "A parent-directed language intervention for children of low socioeconomic status: A randomized controlled pilot study." *Journal of child language*, 43(2016), 366–406.

Tolmer, Zazie. "Qualitative Evaluation Report - TipsByText." Technical report (2021).

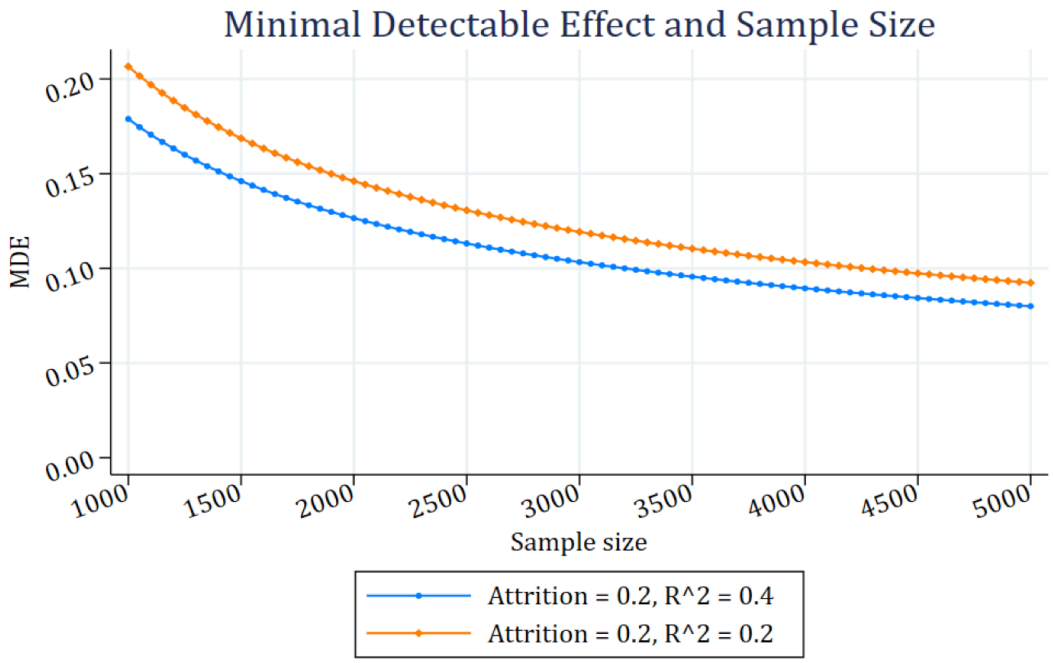
Villadsen, Jakob Waag and Pernille Hviid. "'The history of children's engagements in Danish child care." In "Nordic social pedagogical approach to early years," International perspectives on early childhood education and development. Springer, Cham (2017). 43–62.

York, Benjamin N., Susanna Loeb, and Christopher Doss. "One step at a time the effects of an early literacy text-messaging program for parents of preschoolers." *Journal of Human Resources*, 54(2019), 537–566.

A

Additional Analyses

Figure A.1: Ex ante power simulations



Notes: The figure shows the minimum detectable effect size (MDE) with a power of 80% and a significance level of 5% for different sample sizes given an assumed attrition/opt-out rate of 20% and two different levels of R^2: 0.2, which ex ante was assumed to be the "worst case" scenario, and 0.4, which ex ante was assumed to be the "best case" scenario. The assumed non-compliance (parents who don't follow the intervention's tips but does not opt out) is 10%, the assumed share of siblings is 9%, the degrees of is 1,000, and the assumed standard deviation of the outcome is 1.

Table A.1: Representativeness the of sample within the collaborating municipalities

	Non-sample mean (SD)	Sample difference (SE)	N Sample	N Non-Sample
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: This table shows the means of different variables among parents of children born in years 2014–2016 in the collaborating municipalities (column 1), and the results of a linear regression that tests whether the mean of the parents of children in the study sample differs from the non-study mean (column 2), clustering standard errors on the level of the family, which is the level of the randomization.

Table A.2: Baseline balance between the control and treatment groups: Stratification variables

	Control mean (SD)	Treatment diff. (SE)	N Treated	N Control
Below median at baseline test	0.501	0.000	1385	1400
	(0.500)	(0.019)		
Estimated endline test version = 3 years	0.032	0.000	1738	1736
	(0.177)	(0.006)		
Estimated endline test version = 4 years	0.328	0.010	1738	1736
	(0.470)	(0.016)		
Estimated endline test version = 5 years	0.322	-0.010	1738	1736
	(0.467)	(0.016)		
Estimated endline test version = School (group)	0.317	0.000	1738	1736
	(0.466)	(0.015)		

Notes: This table shows the means of different variables among the control group (column 1), and the results of a linear regression that tests whether the mean of the treatment group is different from the mean of the control group (column 2), clustering standard errors on the level of the family, which is the level of the randomization.

Table A.3: Differential attrition between the control and treatment groups among the children who attrited because they moved

	Control mean (SD)	Treatment difference (SE)	N Treated	N Control
Baseline: Total % correct	0.486	0.010	111	125
	(0.199)	(0.027)		
Baseline: Total % wrong	0.441	-0.010	111	125
	(0.170)	(0.023)		
Baseline: Total % missing	0.073	0.000	111	125
	(0.133)	(0.017)		
Indicator: Boy	0.404	0.170***	132	136
	(0.493)	(0.060)		
Number of siblings	1.396	-0.160	127	134
	(1.163)	(0.142)		
Indicator: Firstborn child (mother)	0.456	0.070	132	136
	(0.500)	(0.061)		
Indicator: One adult in home	0.149	0.000	127	134
	(0.358)	(0.046)		
Indicator: First/second generation immigrant	0.294	-0.070	132	136
	(0.457)	(0.055)		
Indicator: Dad's education high school or less	0.676	-0.090	132	136
	(0.470)	(0.062)		
Indicator: Mom's education high school or less	0.426	0.020	132	136
	(0.496)	(0.064)		
Father income (thousands)	369.303	-29.300	122	132
	(236.343)	(31.416)		
Mother income (thousands)	218.983	15.730	125	133
	(206.317)	(28.537)		

Notes: This table shows the means of different variables among the control group (column 1) and the results of a linear regression that tests whether the mean of the treatment group is different from the mean of the control group (column 2), clustering standard errors on the level of the family, which is the level of the randomization. Columns 3 and 4 show the number of observations for which data for the given variable is available. Comparisons in the top panel are based on the full sample, and comparisons in the bottom panel are based only on the group of children who was missing an endline language assessment.

Table A.4: Differential attrition between the control and treatment groups among the children who attrited but did not move

	Control mean (SD)	Treatment difference (SE)	N Treated	N Control
Baseline: Total % correct	0.514	0.030	103	81
	(0.229)	(0.034)		
Baseline: Total % wrong	0.407	-0.020	103	81
	(0.195)	(0.029)		
Baseline: Total % missing	0.079	-0.010	103	81
	(0.150)	(0.022)		
Indicator: Boy	0.528	0.010	124	108
	(0.502)	(0.067)		
Number of siblings	1.234	0.030	122	107
	(1.271)	(0.163)		
Indicator: Firstborn child (mother)	0.509	-0.020	124	108
	(0.502)	(0.065)		
Indicator: One adult in home	0.159	0.010	122	107
	(0.367)	(0.050)		
Indicator: First/second generation immigrant	0.278	0.010	124	108
	(0.450)	(0.060)		
Indicator: Dad's education high school or less	0.528	0.070	124	108
	(0.502)	(0.067)		
Indicator: Mom's education high school or less	0.537	-0.040	124	108
	(0.501)	(0.068)		
Father income (thousands)	350.383	5.990	119	103
	(234.171)	(30.053)		
Mother income (thousands)	202.187	21.800	120	106
	(178.304)	(25.137)		

Notes: This table shows the means of different variables among the children who attrited from the control group (column 1) and the results of a linear regression that tests whether the mean of the children who attrited from the treatment group is different from the mean of the control group (column 2), clustering standard errors on the level of the family, which is the level of the randomization. Columns 3 and 4 show the number of observations for which data for the given variable is available.

Table A.5: Baseline balance for the evaluation sample: Language skills

	Control mean (SD)	Treatment diff. (SE)	N Treated	N Control
Baseline: Total % correct	0.483	-0.010	1385	1400
	(0.197)	(0.007)		
Baseline: Total % wrong	0.433	0.010	1385	1400
	(0.171)	(0.006)		
Baseline: Total % missing	0.084	0.000	1385	1400
	(0.130)	(0.004)		
Baseline: Comprehension, % correct	0.656	0.000	1385	1400
	(0.190)	(0.007)		
Baseline: Vocabulary, % correct	0.417	-0.010	1385	1400
	(0.237)	(0.008)		
Baseline: Rhyme detection, % correct	0.472	0.000	1385	1400
	(0.356)	(0.011)		
Baseline: Concepts of print, % correct	0.472	0.000	856	903
	(0.262)	(0.011)		
Baseline: Letter identification, % correct	0.377	-0.040**	406	417
	(0.312)	(0.022)		
Baseline: Deletion, % correct	0.449	-0.020	406	417
	(0.298)	(0.022)		
Baseline: Communication, mean	2.754	-0.010	1365	1375
	(0.655)	(0.022)		
Baseline: Age (years) at language test	4.333	-0.040***	1385	1400
	(0.871)	(0.013)		

Notes: This table shows the means of different variables among the subset of the control group who did not attrit (column 1), and the results of a linear regression that tests whether the mean of the subset of the treatment group who did not attrit is different from the mean of the control group (column 2), clustering standard errors on the level of the family, which is the level of the randomization, and controlling for end-line test version and institution fixed effects, as is done in the main analyses.

Table A.6: Baseline balance for the evaluation sample: Socioeconomic variables

	Control mean (SD)	Treatment diff. (SE)	N Treated	N Control
Indicator: Boy	0.497	0.030	1487	1494
	(0.500)	(0.019)		
Low birth weight (< 2500 g)	0.054	0.000	1425	1433
	(0.227)	(0.009)		
Number of siblings	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)		
Indicator: Mother working part time	0.184	-0.010	1477	1475
	(0.387)	(0.015)		
Indicator: Father 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: This table shows the means of different variables among the subset of the control group who did not attrit (column 1), and the results of a linear regression that tests whether the mean of the subset of the treatment group who did not attrit is different from the mean of the control group (column 2), clustering standard errors on the level of the family, which is the level of the randomization, and controlling for endline test version and institution fixed effects, as is done in the main analyses.

Table A.7: Representativeness of the evaluation sample out of full sample

	Attrited mean (SD)	Evaluation difference (SE)	N evaluation	N attrition
Baseline: Total % correct	0.500	-0.010	2785	603
	(0.209)	(0.009)		
Baseline: Total % wrong	0.422	0.010	2785	603
	(0.177)	(0.008)		
Baseline: Total % missing	0.079	0.000	2785	603
	(0.139)	(0.006)		
Indicator: Boy	0.519	0.010	2981	750
	(0.500)	(0.024)		
Low birth weight (< 2500 g)	0.052	0.000	2858	673
	(0.222)	(0.012)		
Number of siblings	1.267	0.070	2963	739
	(1.174)	(0.052)		
Indicator: Firstborn child (mother)	0.497	-0.060***	2981	750
	(0.500)	(0.023)		
Indicator: One adult in home	0.153	-0.030*	2963	739
	(0.360)	(0.017)		
Indicator: First/second generation immigrant	0.293	-0.040**	2981	750
	(0.456)	(0.019)		
Indicator: Dad's education high school or less	0.592	0.060**	2981	750
	(0.492)	(0.023)		
Indicator: Mom's education high school or less	0.503	0.050**	2981	750
	(0.500)	(0.023)		

Notes: This table shows the means of different variables among the attrited group (column 1) and the results of a linear regression that tests whether the mean of the evaluation sample is different from the mean of the attrited sample (column 2), clustering standard errors on the level of the family, which is the level of the randomization. Columns 3 and 4 show the number of observations for which data for the given variable is available.

Table A.8: Treatment effect on different definitions of the outcome measure

	(1) Standardized index	(2) Proportion correct	(3) Proportion wrong	(4) Proportion missing	(5) Δ Prop. corrrect
Treatment	-0.012	-0.004	0.004	0.001	-0.000
	(0.025)	(0.005)	(0.005)	(0.003)	(0.005)
Mean of treated group	0.001	0.570	0.377	0.053	0.103
S.D. of treated group	1.008	0.208	0.179	0.112	0.148
N treated group	1381	1484	1484	1484	1381
Mean of control group	0.036	0.578	0.370	0.051	0.103
S.D. of control group	0.975	0.207	0.177	0.111	0.146
N control group	1399	1493	1493	1493	1399
Preschool FE	✓	✓	✓	✓	✓
Test version FE	✓	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓	
SE cluster level	Family	Family	Family	Family	Family
Observations	2780	2977	2977	2977	2780
R^2	0.609	0.623	0.525	0.329	0.209

Notes: The table presents the results of a regression of a dummy for a child being in the treatment group on five different variants of the outcome measure: the standardized index (column 1), the proportion items answered correctly (column 2), the proportion items answered wrongly (column 3), and the proportion of missing items (column 4), and the change in the proportion of correctly answered items between the baseline and endline assessment. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.9: Probability for being in different quartiles of endline outcome distribution.

	Dependent variable: Probability of being in different quartiles of endline language distribution			
	(1) 1st quartile	(2) 2nd quartile	(3) 3rd quartile	(4) 4th quartile
Treatment	0.007	0.014	-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	0.391	0.070	0.083	0.304

Notes: The table presents the results of a regression of a dummy for a child being in the first (column 1), second (column 2), third (column 3), and fourth (column 4) quatile of the outcome distribution across the treatment and control groups. * p < 0.10, ** p < 0.05, *** p < 0.01.

Table A.10: Treatment effect for different subgroups

	Dependendent variable: Standardized index of language skills			
	(1) Child scores 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	-0.055	-0.010	0.012
	(0.038)	(0.036)	(0.033)	(0.064)
Preschool FE	✓	✓	✓	✓
Test version FE	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓
SE cluster level	Family	Family	Family	Family
Observations	1387	1598	1927	569
R^2	0.447	0.565	0.548	0.560

Notes: The table presents the results of a regression of the standardized language on a treatment indicator, the standardized baseline measure, and an indicator for whether baseline was missing for the four subgroups among whom we expected to see the largest treatment effects. The number of observations differ between groups because information on parental education and immigration status is missing for some children. * p < 0.10, ** p < 0.05, *** p < 0.01.

Table A.11: Treatment effect of the TipsByText intervention on the treated

	(1) ITT	(2) LATE
Treatment	-0.028	
	(0.026)	
Compliance		-0.033
		(0.030)
Preschool FE	✓	✓
Test version FE	✓	✓
Baseline controls	✓	✓
SE cluster level	Family	Family
Observations	2977	2977
R^2	0.554	0.554

Notes: Column (1) replicates the intent-to-treat main effect presented in Table 1.8. Column (2) presents the result of a treatment-on-the-treated analysis in which the endogenous variable "compliance" is instrumented by the exogenous variable "treatment assignment" using two stage least square estimation. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.12: Differential characteristics of children of parents who opted out vs. where at least one parent stayed enrolled in the intervention

	Enrolled mean (SD)	Unsubscribed difference (SE)	N Unsubscribed	N Enrolled
Baseline: Total % correct	0.468	0.060***	261	1338
	(0.201)	(0.013)		
Baseline: Total % wrong	0.445	-0.040***	261	1338
	(0.175)	(0.011)		
Baseline: Total % missing	0.087	-0.020***	261	1338
	(0.139)	(0.007)		
Indicator: Boy	0.536	-0.030	283	1460
	(0.499)	(0.038)		
Low birth weight (< 2500 g)	0.059	-0.010	279	1375
	(0.236)	(0.017)		
Number of siblings	1.287	-0.040	282	1446
	(1.072)	(0.073)		
Indicator: Firstborn child (mother)	0.449	-0.080**	283	1460
	(0.498)	(0.034)		
Indicator: One adult in home	0.112	-0.010	282	1446
	(0.316)	(0.021)		
Indicator: First/second generation immigrant	0.240	-0.110***	283	1460
	(0.427)	(0.018)		
Indicator: Dad's education high school or less	0.654	-0.010	283	1460
	(0.476)	(0.034)		
Indicator: Mom's education high school or less	0.536	-0.030	283	1460
	(0.499)	(0.037)		

Notes: This table shows the means of different variables among the families who did not opt out (column 1), and the results of a linear regression that tests whether the mean of the group that opted out is different from the mean of the group that did not opt out (column 2), clustering standard errors on the level of the family, which is the level of the randomization, and controlling for endline test version and institution fixed effects, as is done in the main analyses.

Table A.13: Additional categories of parent's free text survey responses.

Negative categories the parents mentioned when asked about the TipsByText intervention	Would recommend TbT (N=141)	Would not recommend TbT (N=16)	Unsure (N=29)
Too repetitive	2	3	0
Too many text messages	6	2	4
Felt talked down to	0	1	0
Bad timing of text messages	5	0	1
Unnatural	0	2	0
Too difficult	1	0	1
Too easy/not relevant	5	4	14
Did the activities already	22	9	11

Notes: The table indicates the number of parents who mentioned a given category in their free text responses. 174 parents gave at least one free response. 123 of these parents would recommend TbT, 15 would not recommend TbT, and 28 didn't know whether they would recommend TbT.

Table A.14: Differential characteristics of families where at least one parent indicated engagement with the TipsByText intervention vs. children of parents who opted out

	Opt out mean (SD)	Engaged difference (SE)	N Engaged	N Opt out
Baseline: Total % correct	0.483	-0.070***	254	254
	(0.257)	(0.023)		
Baseline: Total % wrong	0.342	0.000	254	254
	(0.200)	(0.021)		
Baseline: Total % missing	0.049	0.000	254	254
	(0.103)	(0.009)		
Indicator: Boy	0.508	-0.060	254	254
	(0.501)	(0.053)		
Low birth weight (< 2500 g)	0.047	0.010	254	254
	(0.213)	(0.022)		
Number of siblings	1.193	0.170	254	254
	(0.875)	(0.108)		
Indicator: Firstborn child (mother)	0.382	0.070	254	254
	(0.487)	(0.049)		
Indicator: One adult in home	0.083	0.000	251	254
	(0.276)	(0.028)		
Indicator: First/second generation immigrant	0.047	0.150***	254	254
	(0.213)	(0.031)		
Indicator: Dad's education high school or less	0.642	-0.060	254	254
	(0.480)	(0.051)		
Indicator: Mom's education high school or less	0.496	-0.070	254	254
	(0.501)	(0.053)		

Notes: This table shows the means of different variables among the families who opted out of the TipsByText intervention (column 1), and the results of a linear regression that tests whether the mean of the group that indicated engagement with the intervention is different from the mean of the group that opted out (column 2), clustering standard errors on the level of the family, which is the level of the randomization, and controlling for endline test version and institution fixed effects, as is done in the main analyses.

Table A.15: Differential characteristics of families where at least one parent indicated engagement with the TipsByText intervention vs. children of parents who did not opt out

	Enrolled mean (SD)	Engaged difference (SE)	N Engaged	N Enrolled
Baseline: Total % correct	0.404	-0.020	254	1086
	(0.240)	(0.016)		
Baseline: Total % wrong	0.394	-0.030	254	1086
	(0.219)	(0.016)		
Baseline: Total % missing	0.081	-0.030***	254	1086
	(0.135)	(0.008)		
Indicator: Boy	0.534	-0.060	254	1086
	(0.499)	(0.036)		
Low birth weight (< 2500 g)	0.055	-0.010	254	1086
	(0.229)	(0.015)		
Number of siblings	1.288	0.000	254	1086
	(1.071)	(0.078)		
Indicator: Firstborn child (mother)	0.436	0.020	254	1086
	(0.496)	(0.035)		
Indicator: One adult in home	0.117	-0.010	251	1081
	(0.322)	(0.021)		
Indicator: First/second generation immigrant	0.228	0.030	254	1086
	(0.420)	(0.027)		
Indicator: Dad's education high school or less	0.671	-0.060*	254	1086
	(0.470)	(0.038)		
Indicator: Mom's education high school or less	0.576	-0.120***	254	1086
	(0.494)	(0.038)		

Notes: This table shows the means of different variables among the families who did not opt out of the TipsByText intervention (column 1), and the results of a linear regression that tests whether the mean of the group that indicated engagement with the intervention is different from the mean of the group that did not opt out (column 2), clustering standard errors on the level of the family, which is the level of the randomization, and controlling for endline test version and institution fixed effects, as is done in the main analyses.

Table A.16: Differential characteristics of families where at least one parent indicated engagement with the TipsByText intervention vs. children in the control group

	Control mean (SD)	Engaged difference (SE)	N Engaged	N Control
Baseline: Total % correct	0.453	-0.050***	254	1494
	(0.224)	(0.015)		
Baseline: Total % wrong	0.406	-0.040***	254	1494
	(0.196)	(0.015)		
Baseline: Total % missing	0.079	-0.030***	254	1494
	(0.127)	(0.008)		
Indicator: Boy	0.497	-0.010	254	1494
	(0.500)	(0.035)		
Low birth weight (< 2500 g)	0.050	-0.010	254	1494
	(0.218)	(0.015)		
Number of siblings	1.284	0.020	254	1494
	(1.094)	(0.075)		
Indicator: Firstborn child (mother)	0.442	-0.010	254	1494
	(0.497)	(0.034)		
Indicator: One adult in home	0.123	-0.030	251	1484
	(0.328)	(0.020)		
Indicator: First/second generation immigrant	0.194	0.030	254	1494
	(0.396)	(0.027)		
Indicator: Dad's education high school or less	0.637	-0.040	254	1494
	(0.481)	(0.038)		
Indicator: Mom's education high school or less	0.537	-0.100***	254	1494
	(0.499)	(0.037)		

Notes: This table shows the means of different variables among the families in the control group (column 1), and the results of a linear regression that tests whether the mean of the group that indicated engagement with the intervention is different from the mean of the control group (column 2), clustering standard errors on the level of the family, which is the level of the randomization, and controlling for endline test version and institution fixed effects, as is done in the main analyses.

Table A.17: Differential language growth for the engaged vs. children of parents who opted out

	(1) Prop. correct (baseline)	(2) Prop. correct (baseline)	(3) Prop. correct (endline)	(4) Prop. correct (endline)	(5) Δ Prop. correct	(6) Δ Prop. correct
Indication of engagement	-0.037**	-0.025	0.004	0.000	0.019	0.009
	(0.017)	(0.017)	(0.013)	(0.014)	(0.014)	(0.016)
Mean of engaged	0.478	0.483	0.580	0.583	0.102	0.100
S.D. of engaged	0.187	0.184	0.195	0.195	0.146	0.145
N engaged	204	201	204	201	204	201
Mean of synthetic control	0.550	0.550	0.643	0.643	0.093	0.093
S.D. of synthetic control	0.193	0.194	0.186	0.186	0.150	0.150
N synthethic control	214	213	214	214	214	213
Preschool FE	✓	✓	✓	✓	✓	✓
Test version FE	✓	✓	✓	✓		
Test combination FE					✓	✓
Baseline controls	✓	✓	✓	✓		
Full controls		✓		✓		✓
SE cluster level	Family	Family	Family	Family	Family	Family
Observations	418	414	418	415	418	414
R^2	0.544	0.640	0.729	0.753	0.438	0.474

Notes: The table presents the results of a regression of the difference (Δ) in the proportion of correctly answered questions between endline and baseline (columns 5 and 6) on an indicator for whether at least one parent indicated engagement with the intervention, controlling for test version fixed effects because not all children who are assessed with a given endline material is assessed with the same baseline material. The comparison group is all children of parents where all enrolled parents opted out. For reference, columns (1)–(4) show the the results of a regression of the proportion of correctly answered questions at baseline (columns 1 and 2) and endline (columns 2 and 3) on an indicator for parental engagement with the intervention. All regressions control for the standardized baseline measure, an indicator for whether baseline was missing for different subgroups of interest, and regressions presented in columns (2), (4) and (6) control for a full set of demographic control variables. Columns (5) and (6) The bottom panel shows the mean and S.D. of the two groups in order to interpret the size of the differences relative to the control group variance. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.18: Differential language growth for the engaged vs. children of parents who did not opt out

	(1) Prop. correct (baseline)	(2) Prop. correct (baseline)	(3) Prop. correct (endline)	(4) Prop. correct (endline)	(5) Δ Prop. correct	(6) Δ Prop. correct
Indication of engagement	0.019	0.012	0.019**	0.018*	0.012	0.012
	(0.012)	(0.011)	(0.009)	(0.009)	(0.010)	(0.010)
Mean of engaged	0.477	0.481	0.579	0.582	0.102	0.100
S.D. of engaged	0.188	0.186	0.195	0.195	0.146	0.146
N engaged	208	205	208	205	208	205
Mean of synthetic control	0.460	0.461	0.566	0.567	0.106	0.106
S.D. of synthetic control	0.200	0.200	0.204	0.203	0.149	0.149
N synthethic control	947	943	947	943	947	943
Preschool FE	✓	✓	✓	✓	✓	✓
Test version FE	✓	✓	✓	✓		
Test combination FE					✓	✓
Baseline controls	✓	✓	✓	✓		
Full controls		✓		✓		✓
SE cluster level	Family	Family	Family	Family	Family	Family
Observations	1155	1148	1155	1148	1155	1148
R^2	0.383	0.541	0.686	0.701	0.301	0.350

Notes: The table presents the results of a regression of the difference (Δ) in the proportion of correctly answered questions between endline and baseline (columns 5 and 6) on an indicator for whether at least one parent indicated engagement with the intervention, controlling for test version fixed effects because not all children who are assessed with a given endline material is assessed with the same baseline material. The comparison group is all children of parents who did not opt out. For reference, columns (1)–(4) show the the results of a regression of the proportion of correctly answered questions at baseline (columns 1 and 2) and endline (columns 2 and 3) on an indicator for parental engagement with the intervention. All regressions control for the standardized baseline measure, an indicator for whether baseline was missing for different subgroups of interest, and regressions presented in columns (2), (4) and (6) control for a full set of demographic control variables. Columns (5) and (6) The bottom panel shows the mean and S.D. of the two groups in order to interpret the size of the differences relative to the control group variance. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.19: Differential language growth for the engaged vs. children in the control group

	(1) Prop. correct (baseline)	(2) Prop. correct (baseline)	(3) Prop. correct (endline)	(4) Prop. correct (endline)	(5) Δ Prop. correct	(6) Δ Prop. correct
Indication of engagement	0.011	0.006	0.013	0.010	0.010	0.008
	(0.012)	(0.010)	(0.009)	(0.010)	(0.010)	(0.010)
Mean of engaged	0.477	0.481	0.579	0.582	0.103	0.101
S.D. of engaged	0.188	0.185	0.194	0.194	0.146	0.146
N engaged	209	206	209	206	209	206
Mean of synthetic control	0.483	0.484	0.586	0.587	0.103	0.103
S.D. of synthetic control	0.198	0.197	0.200	0.200	0.146	0.146
N synthethic control	1398	1390	1398	1390	1398	1390
Preschool FE	✓	✓	✓	✓	✓	✓
Test version FE	✓	✓	✓	✓		
Test combination FE					✓	✓
Baseline controls	✓	✓	✓	✓		
Full controls		✓		✓		✓
SE cluster level	Family	Family	Family	Family	Family	Family
Observations	1607	1596	1607	1596	1607	1596
R^2	0.362	0.494	0.675	0.684	0.303	0.326

Notes: The table presents the results of a regression of the difference (Δ) in the proportion of correctly answered questions between endline and baseline (columns 5 and 6) on an indicator for whether at least one parent indicated engagement with the intervention, controlling for test version fixed effects because not all children who are assessed with a given endline material is assessed with the same baseline material. The comparison group is all children in the control group. For reference, columns (1)–(4) show the the results of a regression of the proportion of correctly answered questions at baseline (columns 1 and 2) and endline (columns 2 and 3) on an indicator for parental engagement with the intervention. All regressions control for the standardized baseline measure, an indicator for whether baseline was missing for different subgroups of interest, and regressions presented in columns (2), (4) and (6) control for a full set of demographic control variables. The bottom panel shows the mean and S.D. of the two groups in order to interpret the size of the differences relative to the control group variance. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.20: Treatment effect by child age

	Dependent variable: Standardized index of language skills				
	(1) 3-year-olds	(2) 4-year-olds	(3) 5-year-olds	(4) 6-year-olds	(5) School
Treatment	-0.406**	-0.009	0.002	0.104	-0.090
	(0.157)	(0.042)	(0.039)	(0.099)	(0.063)
Preschool FE	✓	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓	✓
SE cluster level	Family	Family	Family	Family	Family
Observations	85	1039	884	266	659
R^2	0.694	0.614	0.707	0.431	0.480

Notes: The table presents the results of a regression of the standardized language on a treatment indicator, the standardized baseline measure, and an indicator for whether baseline was missing for different endline test versions of the language assessment material * p < 0.10, ** p < 0.05, *** p < 0.01.

Table A.21: Treatment effect by baseline language skills

	Dependent variable: Standardized index of language skills				
	(1) 1st quartile	(2) 2nd quartile	(3) 3rd quartile	(4) 4th quartile	(5) Below median
Treatment	0.075	-0.101*	-0.066	-0.018	0.016
	(0.055)	(0.060)	(0.049)	(0.047)	(0.038)
Preschool FE	✓	✓	✓	✓	✓
Test version FE	✓	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓	✓
SE cluster level	Family	Family	Family	Family	Family
Observations	736	806	674	727	1387
R^2	0.383	0.299	0.296	0.355	0.447

Notes: The table presents the results of a regression of the standardized language on a treatment indicator, the standardized baseline measure, and an indicator for whether baseline was missing for different quartiles of the distribution of the baseline language assessment. Quartiles are defined within test version. * p < 0.10, ** p < 0.05, *** p < 0.01.

Table A.22: Treatment effect by parental labor market participation

	Dependendent variable: Standardized index of language skills			
	(1)	(2)	(3)	(4)
Treatment	-0.03	-0.03	-0.01	-0.03
	(0.03)	(0.03)	(0.03)	(0.03)
Mother is outside the labor force	-0.15*			
	(0.09)			
Mother is outside the labor force × Treatment	-0.10			
	(0.13)			
Father is outside the labor force		-0.07		
		(0.14)		
Father is outside the labor force × Treatment		-0.03		
		(0.20)		
Mother is working part time			0.09**	
			(0.04)	
Mother is working part time × Treatment			-0.11	
			(0.07)	
Father is working part time				-0.08
				(0.08)
Father is working part time × Treatment				0.06
				(0.11)
Preschool FE	✓	✓	✓	✓
Test version FE	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓
SE cluster level	Family	Family	Family	Family
Observations	2977	2977	2977	2977
R^2	0.556	0.554	0.555	0.554

Notes: The table presents the results of a regression of the standardized language index on a treatment indicator, an indicator for a given labor market participation, an interaction between the treatment indicator and parental labor market participation, the standardized baseline measure, and an indicator for whether baseline was missing. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B

Language Assessment Details

There are five different versions of the language assessment tool, "Sprogvurdering 3-6;" the versions differ in terms of subtests and questions asked within each subtest (see Table 1.4). Each child is assessed with the version of the tool that corresponds with that child's age (Børne- og Undervisningsministeriet, 2019). Across the five versions, there are eight different subtests, the content of which are explained below.

Comprehension

The subtest "comprehension," which comprises 20 questions, tests the child's language comprehension. For each question, the child is shown four images. The pedagogue asks a question, for example, "Where is a boy being chased by a dog?" and the child has to point to the correct image. For each question, the pedagogue notes whether the child answered correctly, wrongly, or did not answer the question.

In the class-based version of this subtest, each child is given a sheet of paper with 20 rows of four images. Each row corresponds to one question. For each question, the teacher reads the question out loud and the children have to draw a circle around the correct image on their individual sheets of paper.

Vocabulary

The subtest "vocabulary," which comprises 25 questions, tests the children's productive vocabulary. For each question, the child is either shown one image or a pair of related images. If the child is shown one image, she has to say what that image depicts. For example, the child is shown a drawing of an octopus, the pedagogue asks "What is that?" and the child has to answer "octopus." If the child is shown a pair of images, the pedagogue explains what is on the first image, and the child has to explain what is one the other image. For example, the child is shown an empty bottle and full bottle and the pedagogue says "One bottle is empty, what is the other one?" and the child has to answer "full" to get a correct answer. For each question, the pedagogue notes whether the child answered correctly, wrongly, or did not answer the question.

This subtest is not included in the class-based version of the assessment.

Rhyme detection

The subtest "rhyme detection," which comprises 15 items, tests the child's phonological awareness, as the child has to find words that sound alike. For each question, the child is shown three images. The pedagogue points to each of the three pictures in turn and says what is on them, and then asks the child which of the two words rhyme with each other. For example, the child is shown pictures of a house, a mouse and a sun, and the child has to point on the house and the mouse to get a correct answer. For each question, the pedagogue notes whether the child answered correctly, wrongly, or did not answer the question.

In the class-based version of this subtest, each child is given a sheet of paper with 15 rows of three images. Each row corresponds to one question. For each question, the teacher reads the three words out loud, and the children have to draw a circle around the two words that rhyme on their individual sheets of paper.

Concepts of print

The subtest "concepts of print," which comprises 16 questions, tests the child's understanding of concepts of written language. The pedagogue and the child sit together with a book, and for each question, the child either has to answer or point at something in or on the book. For example, the pedagogue asks the question "Show me the first word on this page," and the child has to point on the first word on the page to get a correct answer. For each question, the pedagogue notes whether the child answered correctly, wrongly, or did not answer the question.

This subtest is not included in the class-based version of the assessment.

Letter identification

The subtest "letter identification," which comprises 12 questions, tests the child's knowledge of the Danish alphabet. For each question, the child is shown one letter, and she has to say that letter out loud to get a correct answer. For each question, the pedagogue notes whether the child answered correctly, wrongly, or did not answer the question.

The class-based version of the subtest comprises 28 questions (one for each letter in the Danish alphabet). Each child is given a sheet of paper with 28 rows of four letters. Each row corresponds to one question. For each question, the teacher reads the letter out loud and the child has to draw a circle around the correct letter.

Deletion

The subtest "deletion," which comprises 15 questions, tests the child's phonological awareness and ability to distinguish between words, syllables, and sounds. For each question, the child has to say how a word would sound when a part (word, syllable, or sound) is removed. For example, the pedagogue says "Say 'ballet' without 'ba'" and the child has to say "let" to get a correct answer.

This subtest is not included in the class-based version of the assessment.

Auditory discrimination

The subtest "auditory discrimination," which comprises 20 questions, also tests the child's phonological awareness, as the child has to isolate the first sound in a word. This subtest is only a part of the class-based version of the assessment. Each child is given a sheet of paper with 20 rows of five images. Each row corresponds to one question. For each question, the

teacher reads aloud a question and the children have to circle the correct image on their individual sheets of paper. For example, the teacher asks "What word begins with the same sound as 'mmmouse?' 'Mmmouth,' 'house,' 'cat,' or 'comb?'"

Communication

The subtest "communication," which comprises 15 questions, tests the child's general communication skills. "Communication" is the only subtest that the pedagogue fills out without the child present. It is based on the pedagogue's knowledge of the child's communicative strategies in different situations. For each question, the pedagogue has to answer how often ("never," "rarely," "often," or "always") the child uses the given communicative strategy. "Never" is scored with 1 point, and "always" is scored with 4 points. An example of a question is "An adult reads a story to the child. The child talks and asks questions about the pictures/the story."