



HAL
open science

The effect of Crianza Positiva e-messaging program on adult-child language interactions

Ana Balsa, Florencia López Boo, Juanita Bloomfield, Alejandrina Cristia, Alejandro Cid, María de La Paz Ferro, Rosario Valdés, María del Luján González

► To cite this version:

Ana Balsa, Florencia López Boo, Juanita Bloomfield, Alejandrina Cristia, Alejandro Cid, et al.. The effect of Crianza Positiva e-messaging program on adult-child language interactions. Behavioural Public Policy, 2021, 10.1017/bpp.2021.25 . hal-03498848

HAL Id: hal-03498848

<https://hal.science/hal-03498848>

Submitted on 21 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The effect of *Crianza Positiva* e-messaging program on adult-child language interactions

Balsa, Ana¹, Florencia López Boo², Juanita Bloomfield¹, Alejandrina Cristia³, Alejandro Cid¹, María de la Paz Ferro², Rosario Valdés⁴, María del Luján González⁵

April 2020

Abstract

We assess the effects of the *Crianza Positiva* text and audio e-messaging program on caregiver-child language interaction patterns. The program is a six-month-long intervention for families with children aged 0-2 aimed at strengthening parental competences. Its design exploits behavioral tools such as reminders, suggestions of action, and messages of encouragement to reinforce and sustain positive parenting practices. Families in 24 early childhood centers in Uruguay that completed an 8-weeks workshop were randomized into receiving or not receiving mobile messages. After the program, we videotaped 10-minutes-sessions of free play between the caregiver and the child, and decoded language patterns using automated techniques. The intervention was successful at improving the quality of parental vocalizations, as measured by the parent's pitch range. We also found suggestive evidence of increases in the duration of adult vocalizations. Results are consistent with more frequent parental self-reported involvement in reading, telling stories and describing things to the child. Regarding the child, we find a non-robust decrease in the duration of vocalizations, which we attribute to a crowding-out effect by the caregiver in the context of a fixed 10-minute suggested activity and a more proactive parental role.

Keywords: e-messages, parenting, behavioral economics, language, early childhood

JEL Codes: C93, D91, I20

¹ Department of Economics, Universidad de Montevideo. Prudencio de Pena 2544, Montevideo, Uruguay; phone: +59827074461. Ana Balsa: abalsa@um.edu.uy (corresponding author) Juanita Bloomfield: j.bloomfield@um.edu.uy; Alejandro Cid: acid@um.edu.uy. ² Department of Social Protection, Inter-American Development Bank. Washington, DC. Florencia López Boo: florencial@iadb.org. Ma de la Paz Ferro: mafer@iadb.org. ³ Laboratoire de Sciences Cognitives et Psycholinguistique, Département d'études cognitives, ENS, EHESS, CNRS, PSL University, Paris, France. alecristia@gmail.com. ⁴ Fundación Fe y Alegría, Montevideo, Uruguay. crianzapositiva.rosario@gmail.com. ⁵ Department of Psychology, Universidad Católica del Uruguay, Montevideo, Uruguay; mgonzalezturnaria@gmail.com.

We acknowledge financial support from Reaching U Foundation, the Inter-American Development Bank, and Agencia Nacional de Investigación e Innovación (ANII, Uruguay). Alejandrina Cristia acknowledges Agence Nationale de la Recherche (ANR-17-CE28-0007 LangAge, ANR-16-DATA-0004 ACLEW, ANR-14-CE30-0003 MechELex, ANR-17-EURE-0017); and the J. S. McDonnell Foundation. The RCT design for this project was registered in the AEA RCT registry with RCT ID number AEARCTR-0003585. The project was approved by the Institutional Ethics Review Board at Universidad Católica del Uruguay, resolution A 22-08-17.

1. Introduction

Wide variability in language development is present from very early in life, as early as 8 months (Fenson et al., 1994). Early experiences are among the most important predictors of language development, along with maternal education. Research has found that parental interventions that target caregiver's responsiveness and communication patterns are effective in changing parental behavior and promoting language development, with some reports of long-term effects on academic achievement and economic outcomes in adulthood (Gertler et al, 2014). However, these programs are in many cases resource-intensive and difficult to scale up.

In this paper, we evaluate the effects of a low-cost e-messaging intervention directed at parents with children below the age of 2 on the caregiver-child dyad communication patterns. The *Crianza Positiva* e-messaging intervention consists of a series of text and audio messages delivered to the mobile phones of families three times a week for a period of six months. The messages were designed to complement a parenting workshop of eight weekly sessions. Both the workshop and the messages are organized around four parental competences: attachment bond, protection, stimulation, and reflective function. The messages include reminders, suggestions of simple actions, and messages of encouragement to reinforce and sustain positive parenting practices. Behavioral economics tools have demonstrated large benefits in diverse areas of public policy and are increasingly being used in policies targeting families with children (Gennetian et al, 2016; Mayer, Kalil et al, 2018; Ajzenman and Lopez Boo, 2019).

The e-messaging intervention was randomized across 24 early childhood centers in Uruguay that had previously implemented the *Crianza Positiva* workshop (529 families). All treated and control families participated in the workshop, but only those in the treatment received mobile messages. Bloomfield, Balsa, and Cid (2020) analyze the impact of the *Crianza Positiva* e-messaging program on parental self-reported investment and find that the intervention increased the frequency of parental involvement with the child in social, stimulating, and physical activities by 0.24 standard deviations. It also increased parents' investment quality as measured by a positive parenting index and by an index of outreach for social support.

In this paper, we further contribute to the assessment of the program by relying on objective, rather than self-reported, outcomes to study parent-child language interactions. Language development is an important part of the *Crianza Positiva* curriculum: about 20% of the messages focus explicitly on promoting parents' involvement in early literacy practices, while the rest seek to improve parental sensitivity and responsiveness, household environment, and the quality of the caregiver-child interactions, all features indirectly associated with better communication patterns and more successful language development (Aboud et al., 2013). Our analysis of language interactions relies on the automated decoding of a 10-minutes-session of free play between the caregiver and the child,

videotaped around three months after the e-messaging intervention ended. The use of externally assessed outcomes mitigates usual problems of self-reports, such as desirability bias.

Our findings show that the intervention was successful at influencing parental language cues: it increased the duration of caregiver vocalizations and improved the quality of vocalizations with an increase of the adult's pitch range. We did not find clear evidence of positive benefits in child outcomes, which may suggest limitations in short free play sessions for assessing language and cognitive development.

Our work contributes to a nascent literature exploiting audio and text-based communication and behavioral economic insights to boost early literacy development. We believe this is one of the first papers to assess the effectiveness of a large-scale e-messaging intervention using automatic decoding of videotapes to infer language and communication patterns. Our results demonstrate the potential of these very low cost interventions, based on mobile technology and on the understanding of behavioral biases, for changing parental behavioral patterns and ultimately improving child development, particularly when interventions are conducted as a complement to an initial face-to-face intervention with the families.

The paper is structured as follows. Section 2 provides background literature on interventions that promote language development. Section 3 describes the *Crianza Positiva* messaging intervention. Section 4 provides an overview of the methodology used. Section 5 reports our results and in Section 6 we discuss our findings.

2. Background

2.1 Language development in early childhood

The importance of language development during early childhood is well known. It predicts the latter results during youth, including literacy skills and school success (Zimmerman et al., 2009; Rowe, Raudenbush and Goldin-Meadow, 2012). It has also been associated with adulthood outcomes, with significant correlations between language difficulties early in life and lower literacy, poor mental health, and higher unemployment rates (Law et al., 2009). A majority of this research has been carried out in North America, but the little evidence that exists from Latin American samples seems to align with the global trends we discuss next (e.g., Law et al., 2009 in Jamaica; Shneidman & Goldin-Meadow, 2012 in Mexico).

Language development gaps appear early in life. One of the most important predictors of language development are early experiences, along with maternal education. Differences in the rate of language development depend on access to language and gesture from caregivers, and on infant's speech processing skills (Weisleder & Fernald, 2013). Language inputs and social interactions support child's

language development through opportunities where the child can interpret language and practice his/her emerging skills.

Studies have found a positive association between the quantity of language input and the child's vocabulary and speed of word recognition (Hurtado, Marchman & Fernald, 2008; Fernald, Marchman & Weisleder, 2013; Weisleder & Fernald, 2013; Mendive et al., 2020; Shneidman & Goldin-Meadow, 2012). Furthermore, when studying how the two above-mentioned factors work together, Weisleder and Fernald (2013) find that the effect of child-directed speech quantity (language input) on vocabulary is mediated by infants' language processing efficiency. Regarding language input, the evidence points to the conclusion that what matters most for language learning is infant-directed speech and responses contingent on their communicative signals or turn-taking (Snow & Ferguson, 1977; Weisleder & Fernald, 2013; Bornstein et al., 2015; Romeo et al., 2018), rather than speech overheard by the child (Weisleder & Fernald, 2013) or the sheer number of words (Romeo et al., 2018). Caregiver infant-directed speech has been associated with increased depth of processing in brain imaging studies (Zangl & Mills, 2007) and improved early language skills (Weisleder & Fernald, 2013; McGillion et al., 2017). In addition, studies have reported a positive association between the number of conversational turns, child language development (Zimmerman et al., 2009) and brain activity (Romeo et al., 2018). A recent study by Rubio-Codina and Grantham-Mc Gregor (2020) shows that the Family Care Indicator (FCI), an index measuring the quality of the home environment in the first years of life, which assesses, among several other resources and caregiving activities, reading, looking at picture books, telling stories, and singing to the child, can predict IQ and academic performance when the child is in elementary school.¹

Language input has been linked to household socioeconomic status (SES). For the US, Hart and Risley (1995) found that children in low-income households not only heard fewer words per hour from their caregiver than children in high-income households, but the content also varied greatly between groups. The difference they found meant that by age four, children in high-income families will have been exposed to 30 million more words than children from low-income families. A more recent study found that associations between vocabulary and language processing efficiency, and SES, appear as early as 18 months, and that this association is pronounced by the time the child turns two (Fernald, Marchman & Weisleder, 2013). In Latin America, language disparities widen with age and are also associated with SES (Berlinski and Schady, 2015). There is also evidence from Uruguay on SES gradients on the home environment and the FCI (Lopez Boo et al., 2018).

¹ Most of the referenced studies analyze populations from developed countries (United States or United Kingdom) of varied socioeconomic status. Some exceptions are Snow & Ferguson, 1977, that analyze populations from mostly developed and a few less-developed countries; Mendive et al., 2020, that study a varied SES population in Chile; and Rubio-Codina and Grantham-Mc Gregor (2020), who work with a low and middle income population in Colombia.

Other authors have reported large variance in language input and infants' language development within households of homogeneous socioeconomic characteristics (Weisleder & Fernald, 2013), as well as associations over and above SES (Zimmerman et al., 2009; Romeo et al., 2018), suggesting that SES by itself does not determine the infant language trajectory. Sperry, Sperry and Miller (2019) test Hart and Risley (1995)'s (HR) findings using data from five communities in the United States. Their results do not support HR findings: they do not find a clear pattern between number of words spoken by the primary caregiver to the child and SES. Even when testing for a broader definition of the child's verbal environment (all speech directed to the child and speech heard by the child), they still do not find a consistent relation between these measures and SES.

Either way, creating language-rich experiences for children is extremely important to prevent language disparities that continue throughout the life cycle. Parental interventions that target caregiver's responsiveness and communication patterns have been found to be effective. For example, a six-month home visiting program (3Ts: Tune In, Talk More and Take Turns with the child) for low SES households in the United States that focuses on the home language environment, had immediate positive effects on caregivers' knowledge, the quantity of linguistic input and the quality of caregiver interactions (Leung, Hernandez & Suskind, 2020). Some aspects of the intervention the researchers point out as being key to effectively changing parents' behavior at home are: video modeling (strategies are presented through true-to-life videos), concrete feedback (based on quantitative measures of linguistic input from LENA² recordings), and goal setting (goals are set in terms of the linguistic input measures collected from the previous week).

McGillion et al., (2017) evaluate another language intervention that focuses on contingent talk and find immediate effects on caregivers. Households are visited twice in a two-week period. In the first visit, the importance of contingent talk is introduced, a short video with examples is presented, and caregivers are asked to set aside 15 minutes per day to practice what was learned. In the second visit, further questions are solved. Immediately after the end of the intervention, the study finds that parents assigned to the intervention engage in more contingent talk. However, a year later no effects are found. This fade-out suggests the importance of guiding and providing the caregivers with more tools to effectively support the infant's language learning through the different stages of the child's development.

A recent set of studies evaluates the incorporation of technology on early childhood interventions that promote literacy skills. Doss et al., (2019) evaluate "Ready for K!" a text-based intervention in the United State to help parents support their preschoolers' development. The program increased parental involvement at home and at the school and improved child literacy skills. Doss et al., (2019) show that

² Language ENvironment Analysis (LENA) System is a relatively new recording technology that can be used to investigate typical child language acquisition. More details can be found here: <https://www.lena.org/>

personalizing “Ready for K!” messages increases the likelihood that parents read to their children by 50%. Meuwissen et al., (2017) and Hurwitz et al., (2015) also report that the use of text messages increased parental engagement in literacy and other learning activities with their children. Mayer et al., (2018) evaluate an intervention that provided parents with a tablet containing children’s books and a goal setting tool, reminders and social gratification messages. The authors find that parents in the treatment group used the tablet more and read more to their children.

2.2 Measurement of language input and output

In the economics literature, the norm has been to use standardized tests, such as Bayley Scales (Bayley, 2006), the Denver Developmental Screening Test (Frankenburg et al., 1990) and MacArthur Communicative Development Inventory (Fenson et al., 1993), to measure early language development. We know these measures are predictive of future outcomes (Fernald et al., 2009), and that early stimulation strongly impacts language as measured by these tests (NICHD, 2000, Rubio-Codina and Grantham-Mc-Gregor, 2020). However, these measures are subject to limitations including the fact that they are expensive to implement, depend on evaluators' training and profile, and measure the child’s development in a particular setting. An alternative to standardized development tests are audio recordings that measure the child’s language skills in a natural setting in addition to collecting other measures, such as maternal vocabulary and conversational turns. These audio recordings are usually obtained without the mediation of a trained facilitator and provide critical information for understanding the infant’s language trajectory and environment.

There are broadly two methods to capture both quantity and quality of language input and output through recording, namely short audio- or video-recordings, or day-long audio-recordings (Lavechin et al., submitted). In the psycholinguistic literature, it is most common for the former to be transcribed by trained annotators, and for the latter to be analyzed automatically, typically using speech technology algorithms that do not produce orthographic transcriptions (not technically possible today), but which instead provide estimates of who spoke when and for how long (e.g., Zimmerman et al., 2009). Human transcription of adult speech in such conversations is a time-consuming process, with most annotators requiring 20 times the recording duration — i.e., 1 minute of interaction takes 20 minutes to transcribe. Transcription is even more time consuming for speech by children aged 1-3 years, with estimates for transcription time reaching 40 times the recording duration. It is beyond the scope of this article to fully lay out the advantages and disadvantages of short versus long recordings, but we do point out that day-long recordings remain quite challenging to collect in large-scale projects like ours for logistical, legal, cost, and ethical reasons, which led us to prioritize short observations. We believe our study is among

a few to analyze the impact of a behavioral and parenting intervention at scale, and measure language outcomes using short videotaped recordings (a cost-effective method vis à vis day-long recordings).³

In short video or audio recorded observations, quantity of caregiver speech can be measured in terms of number of words, number of vocalizations, or amount of time spoken. We are not aware of research showing that longer vocalization durations correlate with child development in short observations, but in transcription-based research, children who are more advanced produce more vocalizations (Gilkerson et al., 2017) and utterances with more words (i.e., with higher “mean length of utterance”, e.g., Allen & Dench, 2015), which, all else equal, should be correlated with greater child vocalization duration. We also expect more parental input (larger duration of parental vocalizations) to have a positive impact on the child (e.g., Ferjan-Ramírez et al., 2019, 2020).

Communication quality, on the other hand, may be captured through the number of conversational turns, parental responsiveness, and caregiver’s pitch. A child is more likely to be engaged when the conversation takes place in turns rather than when the caregiver talks non-stop. There is research also suggesting that more responsive caregivers, who vocalize after the child vocalized, lead to better language outcomes in day-long recording data (Ferjan-Ramírez et al., 2020). Another way to engage the child is via producing speech that captures the child’s attention. This is often done by using a child-directed speech register (or Parentese speech), characterized by a higher pitch (and in some cultures, larger pitch range) than adult-directed speech. It is in part because of the higher pitch and larger pitch range that infants show a strong preference for infant- over adult-directed speech (Fernald & Kuhl 1987; The ManyBabies Consortium 2020), and this heightened attention boosts performance in a wide range of tasks (a systematic review in Cristia 2013). Caregivers can be trained to use this register, leading to positive child outcomes (Ferjan-Ramírez et al., 2019, 2020).

3. The *Crianza Positiva* Text and Audio Messaging Intervention

The *Crianza Positiva* text and audio program is a six-month-long intervention for families with children aged 0-2 years. Messages were designed to help caregivers change their decision-making strategies in order to adopt what previous research suggests to be positive parenting practices and sustain these practices over time, to ultimately change their children’s development. Participant families started receiving messages right after completing a workshop of eight weekly sessions that was organized around four parenting competences: attachment bond, stimulation, protection, and reflective function. Families in the treatment group received messages three times a week both in text (via SMS) and in

³ Some previous evaluations of parenting programs such as Rodriguez et al., (2010) for Health Families New York, Love et al. (2005) for Early Head Start, and Olds et al., (2004) for Nurse Family Partnership have used measurements based on video-recordings to analyze mother-child interactions, which could thus provide a view both on input and child output. More recently, Conti et al., (2020) also used 3-minutes videotaped play sessions to evaluate a home-visiting program in Germany and focused on reciprocity and initiation of interactions.

audio format (via WhatsApp). The contents of the text and audio messages were exactly the same, except for the fact that text messages were personalized with the name and gender of the child. Control families received a single SMS message thanking them for their participation in the workshop.

The message structure is based on behavioral economics insights. Its design recognizes the existence of four key behavioral biases associated with low parental investment. First, parents' behavior may be influenced by the *present bias* (Thaler, 2015) which implies attaching low valuation to costly activities that imply rewards only in the future. Another key bias is *limited attention and cognitive fatigue* as it deviates cognitive resources from important decisions (Schilbach, Schofield & Mullainathan, 2016; Mani et al., 2013; Shah, Mullainathan & Shafir, 2012) such as those relating to parenting. Lastly, *negative identities* can lead to sub-optimal parental investment by making parents attach a low value to their existing resources and individual performance. Having a positive identity as a parent is important because self-esteem can increase effort through higher levels of motivation (Bénabou & Tirole, 2002).

As stated above, the literature on language development highlights the importance of exposing the child to language stimulation in the context of a sensitive and caring interaction. Throughout the program, several modules of messages focus on providing the right setting for an adequate development of literacy skills. A first module promotes parental sensitive observation, interpretation, and response to the child's needs and signals, gentle treatment, and the expression of affection (17% of the messages). This module emphasizes observing the child carefully, trying to interpret his/her expressions, gestures, and signals, seeking to understand what the child is feeling and thinking, exploring how to get closer to the child by responding to these signals, demonstrating love, and communicating assertively. A second module explicitly emphasizes the benefits of parental involvement in specific home-literary practices (17% of messages), and suggests concrete activities, such as engaging in conversations with the child, asking open questions, describing objects, places and activities, singing and listening to songs together, and engaging in book-reading and story-telling routines. The rest of the modules cover other parenting practices such as child safety at home and the incorporation of routines (17% of messages), free play and parental involvement in enjoyable activities with the child (17%), and parental self-care and reflective function (32%). To the extent that all messages influence the quality of the environment and of the parent-child interaction, they are directly or indirectly setting the grounds for better communication patterns and improved language development.

Treated families receive messages three times a week at the same time (specifically, Mondays, Tuesdays and Fridays at 6pm). A fundamental challenge of home visits and other interventions seeking to improve parenting practices is to be able to help parents overcome potential behavioral barriers they face at home in key moments when parenting decisions are critical. The message structure is designed to address some of these biases and always follows the same pattern. On Mondays, families receive a message with information on the benefits to the child of engaging in a certain parenting practice. This

message aims at addressing present bias and inattention. Reminders have been successfully implemented in many contexts (from reminding people about voting to reducing missed appointments in hospitals or increasing medication adherence) and work well especially if the timing is right (Dale and Strauss 2009; Gurol-Urganci et al., 2013).

On Tuesdays, families receive a suggestion to engage in a specific activity with the child in relation to that week's topic and to repeat it during the week. This activity does not take much time and is usually concrete and easy to implement. The purpose of this message is to simplify parental decision-making in the face of the complexity of parenting, and to help parents establish routines by engaging repetitively in the mentioned practices. The message addresses inattention and cognitive fatigue. Doss et al., (2019) show that providing activity tips to parents can help reduce parental cognitive load and improve parental ability to process information, in particular in the context of highly stressful environments.

On Fridays, families are invited to reflect on their performance during the week and on their personal feelings regarding the task proposed, and usually receive a motivational or encouragement message. This message seeks to help parents overcome negative identities and strengthen self-efficacy.

For example, during the week that focused on vocabulary development and the importance of speaking to the child, parents received the following messages:

MONDAY:

Children's brains are like sponges, they absorb everything: the sounds, the pitches of voices, the language they listen to. The more words your child listens to at this stage, the more [she] will develop [her] language. It is therefore very important that you speak to [child's name], this will impact heavily on [her] ability to learn.

TUESDAY:

Take advantage of every moment to tell [child's name] what you're doing when you're with [her]. For example, say "I'm changing your diapers", "Now I'm preparing the food", or "Let's go for a walk". This will help enrich [her] language.

FRIDAY:

The more you speak to [child's name], the better will [her] language develop and the more [she] will learn. Today and in the following days, remember and repeat this thought: "I take advantage of all the moments with [child's name] to speak to [her]."

The program included some additional components that provided parents with further sources of information and material. For example, we coupled messages promoting reading and singing to the

child with a message that provided a username and password to access "Radio Butiá", a Uruguayan web server that hosts Latin American songs and stories online. These messages addressed concerns relating to the possibility that families without children’s books and music would not perform the suggested activities.

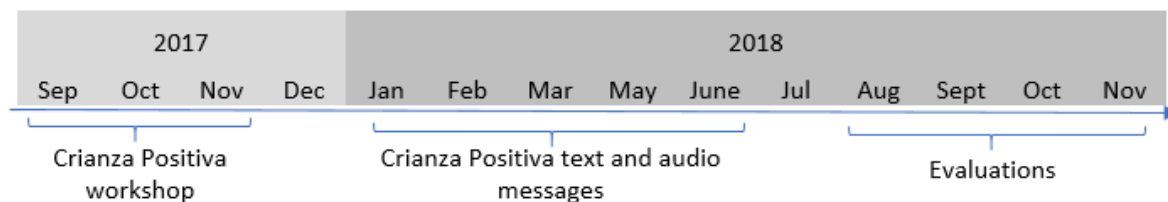
4. Methodology

4.1 *Timeline and main features of program implementation and evaluation*

Participant families were selected from a pool of candidates that had attended an 8-week parenting workshop at “Children and Family Care Centers” (CAIF) of Uruguay. After workshop completion, in December 2017, we assigned families to a treatment and control group to evaluate the effects of the *Crianza Positiva* text and audio messages component. Following Baird et al., (2018), we selected a two-stage randomization strategy in order to measure spillovers. First, we stratified the sample by average maternal education and, within each stratum randomly allocated early childhood centers to receive or not receive messages using a 60%-40% rule. As a result of the randomization, we assigned 14 early childhood centers to the treatment group and 10 centers to the control group. Next, within each center in the treatment arm, we randomized families into receiving or not receiving messages. In centers that were treated, 80% of families were selected to receive messages. Out of the initial 529 families, 237 families were assigned to receive messages, 59 were assigned not to receive messages but belonged to a treated center, and 233 families were assigned to a pure control group.

Figure 1 illustrates the timeline of the implementation and evaluation of the *Crianza Positiva messaging program*. Treated families received messages between January 5, 2018 and June 27, 2018 and Control families received a single SMS message on January 26, 2018 thanking them for participating in the parenting workshop.

Figure 1: *Crianza Positiva* Timeline



Messages were sent to all contact numbers we had for each family (mothers, fathers and other caregivers), which in total added to 373 mobile phones. SMS (text) messages were sent through a platform that enabled us to personalize the messages by including the name of the baby and gender-

specific pronouns and sending them to families at the same time. We created a WhatsApp broadcast list to deliver audio messages.⁴

We collected data on families' socioeconomic status and other characteristics at three points in time. Baseline information was collected before the initiation of the workshops (in August 2017), and after the workshop ended (November-December 2017). We conducted a follow-up assessment two-months after the messaging intervention ended (between August and November 2018). Baseline evaluations were carried out at the CAIF center and follow-up evaluations were carried out either at the family's home or at the center depending on the family's preferences⁵. In this paper, we focus on a specific component of the family assessment: a 10-minute video recording of the interaction between the caregiver and the child in a free play activity. Videos were recorded using a mobile phone either at the family's home or a CAIF center. Evaluators (most of them psychologists, therapists or social workers, or advanced students of these specialties) provided the dyad with a set of toys that they could use during the activity. The mother received instructions to play freely while being recorded. The evaluator was instructed not to interfere.

4.2 Outcomes

The first set of outcomes we consider includes the counts and duration of vocalizations for both the child and the adult. A vocalization is a section of audio that has been classified automatically as being speech-like and attributed to an adult or a child.⁶ As summarized in Section 2.2, previous research suggests that better input is associated with more and/or longer child-directed caregiver vocalizations (e.g., Ferjan-Ramírez et al., 2019, 2020), which here should be reflected in higher count and duration of adult vocalization. As for the child, more advanced children produce more vocalizations (Gilkerson et al., 2017) and potentially longer ones (Allen & Dench, 2015).

Second, we capture communication quality through the number of conversational turns between child and adult. There is a turn whenever there is a threshold of at most 0.3 seconds between a vocalization offset and the next vocalization's onset. We also account for parental responsiveness by considering the proportion of child vocalizations that are followed by an adult vocalization in less than 2 seconds (adult response ratio). If the intervention improves communication quality, this may be reflected in higher turn counts and higher adult response ratios in the treatment compared to the controlled group (see Section 2.2).

⁴ Due to specificities of WhatsApp, only mobile phones that saved the mobile number of *Crianza Positiva* received the messages. In our welcome message, we asked families to add our cellphone to their contact list.

⁵ This was a family choice and was unrelated to treatment status.

⁶ The speech detection software is unable to classify overlapping speech.

Finally, we account for two qualitative measures of adult vocalizations: pitch range and pitch average. Fundamental frequency (f_0) range is the difference between the highest and the lowest f_0 frequency reached within a given vocalization. F_0 average is the average pitch of that vocalization. Research summarized in Section 2.2 leads to the prediction that caregivers in the treatment group should exhibit higher pitch and larger pitch ranges than caregivers in the control group.

All outcomes were obtained from automatic analyses of the 10-minute videos using the DiViMe open-source package (Le Franc et al., 2018) via the avutils package (Neumann, 2020). For the main outcomes, we first conducted manual annotations of the outcome for a random set of 30 videos, and then checked the Pearson correlation between the manual and automatic annotations. These correlations are reported in Appendix A. This allowed us to select the DiViMe diarizer (and parameters, such as inter-turn duration) leading to the best performance. We used the Yunitate broad class diarizer, a recurrent neural network that has been trained to classify audio into three broad classes: male adult, female adult, and child. With this diarizer, we find correlations of $r = .78$ between manual and automatic child vocalization counts, of $r = .64$ for child vocalization durations, and of $r = .77$ for conversational turns. The correlations between manual and automatic annotations for adult vocalizations are: $r = .74$ for vocalization counts, $r = .66$ for vocalization duration, and $r = .7$ for adult response rate within two seconds. Correlations are measures of association between two continuous variables; a correlation of $.7$ is very high (given Cohen 1988's guidelines, where correlations above $r = .5$ are "large"), and it indicates that the two continuous variables share 50% of the variance.

Because child language may be difficult to assess at these ages, we correlated child outcomes with age in months. We expected that child measurements should correlate with age, because we would like them to be sensitive to language development and age is a good proxy for development (older children have better language than younger children). Unfortunately, we did not find a statistically significant correlation between count or duration of child vocalizations and child's age. This raises some concerns about our ability to measure children's language outcomes with precision. We did find, however, a negative correlation between the duration of adult vocalizations and children's age ($r = -.11$; $p < .1$), suggesting that adults vocalize less as children vocalize more in the context of the proposed fixed time activity, and that the measures of adult vocalizations may be more reliable than those of the child. We found no association between maternal age or education and language outcomes.

4.3 Compliance

Issues like having the incorrect mobile number or sending messages to a line that is unavailable or has no credit could lead to SMS failure. In our intervention, we sent messages via SMS and WhatsApp to reduce concerns that this could happen. One month after program initiation, we re-contacted all treated families to verify they were receiving the SMS messages. We repeated this for a random sample of treated families 3 months after program initiation. We learnt from these follow-ups that out of the 237 families in the treatment group, 197 were receiving SMSs correctly while 40 were not. We asked CAIF centers to provide us with new contact numbers of these families. We could update nine contact numbers which reduced our SMS failure rate to 13%. WhatsApp messages had a high reception rate: in 87% of families at least one family member received the messages. The percentage of families that read the messages was also high: families that received the messages read on average 69% of them. Overall, we found that 27 out of the 237 families assigned to treatment did not get SMS nor WhatsApp messages. The correlation between assignment to the messaging treatment and receiving the messages was $r=.9$.

4.4 Attrition and balance

We managed to collect videos on 50% of the 529 families at baseline (either in August or November-December 2017) and on 60% of the families at follow-up (317 families overall, 122 assigned to receive messages and 195 in the control group). During baseline, control families were less likely than treated ones to consent to being videotaped, while the opposite was true at follow-up. We ended up with 178 families (34%) with a recorded video both at baseline and at follow-up, 80 of which were in the messaging treatment group and 98 in the control group.⁷ There was no differential attrition in these families across treatment and control arm.

Based on WhatsApp message reception markers, eighty percent of treated families with a video at follow-up read or listened to the WhatsApp messages at least once. And among these, families opened on average 79% of the messages. Regarding families with videos at baseline and follow-up, 85% read or opened at least one WhatsApp message and on average, they opened 83% the messages.

Table 1 reports descriptive statistics of family characteristics at baseline as well as the balance of these characteristics across treated and control subjects (ITT=1 and ITT=0, respectively⁸), both for the sample of families with videos at baseline and follow-up as for the sample only with a video recorded at follow-up. Regarding the first sample (see columns (1) through (3)), more than 90% of caregivers correspond

⁷ 18 of these were control families within treated centers.

⁸ ITT=1 whenever the family was assigned to receive messages. It is 0 when the family participates in a center not assigned to be part of the treatment, or participates in a center selected for treatment, but was not assigned to receive messages.

to the child's mother. The average age of the mother is around 30 years old, nearly half of the children are male, around 11% were born premature, and 20% are still being breastfed. Three out of four children live in a family with both his/her biological parents, on average children live with another sibling in the house and 15% live with other adults. We constructed a wealth index that was a weighted average of household assets, with the weight of each asset in inverse proportion to its availability across households. The index is skewed to the left: in a range of 0 to 1, the average household has an index value of 0.25. Two out of three households receive governmental assistance and 85% report having participated in at least 6 of the 8 *Crianza Positiva* workshop sessions. Twenty-nine percent of mothers have completed middle school, but not high school, and another 28% reports having completed high school or more. This implies that more than 40% of mothers have less than middle school. Forty percent of mothers work full time, and around 20% work part time, and 80% of fathers have a full-time job. Half of the families suffered a negative shock in the past 12 months, such as divorce, a death, a job dismissal, or a disease within the family, among other issues.

Regarding behavioral biases, we use the parent's discount rate to capture present bias, the number of negative shocks faced by the family to measure cognitive fatigue, and a parental efficacy scale to capture negative identities. Specifically, we use Kirby, Petry, and Bickel (1999)'s Monetary Choice Questionnaire (MCQ) to measure the discount rate. Because the number of observations for the MCQ at baseline was quite small (N=115), and we did not expect treatment to affect the parameter, we work with the MCQ as measured at follow up (N=148).⁹ The negative shocks variable reflects the total number of the following events experienced by the family in the 12 months prior to the follow up survey: divorce, separation, moving, income falls, debt, drug problems in the family, the death of a close friend or a family member, unemployment, job problems, or problems with the law. We use the measure at follow up because we do not expect treatment to affect these events and we seek to capture family stress or emotional fatigue during the time the intervention took place. Finally, we use the Johnston and Mash (1989) version of the Parental Efficacy Subscale in the Parental Sense of Competence Scale (PSOC).¹⁰ In this case, we choose to work with the baseline measure because we do expect treatment to affect parental efficacy.

Column (3) in Table 1 checks for balance in family baseline characteristic across treated and control groups. Out of 22 characteristics, we only find a statistically significant difference at 5% or less in the

⁹ The geometric mean of the MCQ discount rate k in our sample ($\ln(0.08)=-2.53$) is larger than that found in Phillips et al. (2019) for a sample of overweight adults aged 18 to 60, and in Kirby et al. (1999) for a sample of opioid dependent patients (3.96). Our measure is within the upper quartile of the k distribution in the cited papers.

¹⁰ The average of the PSOC efficacy subscale in our sample is 4.13, similar to the 4.16 found among Spanish families with children aged 0–5 (Alvarez et al., 2021), but below the average scale in other English speaking samples, such as in Karp et al., 2015 (4.85).

mother's age: treated mothers are 2.6 years older than control mothers. At a 10% level of significance, we also find that treated families have a larger number of children, on average, in the family. The fact that we only find a difference at a 5% level in 1 out of 22 tested features suggests that our sample is balanced.

The sample with videos at follow-up (but not necessarily at baseline) is quite similar in terms of family characteristics to the other sample.¹¹ But unlike the former, it shows differential attrition: treated families are 15 percentage points less likely to be videotaped at follow-up than control families.¹² This results in a slightly higher imbalance of baseline characteristics. At a 5% significance or less, mothers in the treatment group are older than those in the control, and the child is less likely to be an only child. There are more differences at the 10% level, with children in the treatment group being younger and living with a larger number of other children and adults in the household. While only two out of 22 covariates are unbalanced at normal levels of significance, because 5 covariates are unbalanced at a p-level of 0.10, in the next section we devise some strategies to address the differential attrition.

¹¹ Some minor differences include being less likely to have premature babies, being more likely to breastfeed, having more children and more adults in the household, and being more likely to have a negative shock in the past 12 months.

¹² We offered families the option of being videorecorded at the early childhood center or at home. Some centers were more generous than others in offering us physical space and operation hours to record the videos. Because randomization was conducted at the center level, any barriers to videorecording imposed by a single center could substantially affect the balance. The observed imbalance is, in fact, largely explained by one of the centers assigned to treatment, that accounted for more than 10% of the treated families, and in which we were only able to videorecord 8% of the dyads. In all the other centers participation was at least 40%.

Table 1: Descriptive statistics and balance of family characteristics at baseline

	Sample with videos at baseline & follow-up			Sample with videos at follow-up		
	ITT=0	ITT=1	Diff.	ITT=0	ITT=1	Diff.
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Probability of being videotaped</i>	0.34	0.34	0.002	0.67	0.52	-0.153***
<i>Family characteristics</i>						
Caregiver: mother	0.90	0.94	0.040	0.91	0.94	0.030
Mother's age	28.97 (6.81)	31.61 (6.84)	2.643**	28.31 (6.51)	30.60 (6.70)	2.281***
Child's gender	0.47	0.50	0.031	0.49	0.48	-0.017
Child's age in months by Jan 1 2018	25.49 (5.02)	24.76 (6.11)	-0.729	24.44 (6.03)	23.09 (6.67)	-1.344*
Premature	0.13	0.09	-0.047	0.10	0.06	-0.041
Breastfed	0.20	0.19	-0.017	0.26	0.24	-0.023
Lives with both parents	0.78	0.76	-0.016	0.78	0.74	-0.036
Only child	0.41	0.31	-0.099	0.41	0.28	-0.129**
# of other children in the household	0.81 (0.97)	1.12 (1.26)	0.305*	0.93 (1.19)	1.20 (1.25)	0.279*
Other adults in the household	0.17	0.14	-0.027	0.22	0.13	-0.089*
Precarious roofs or floor	0.31	0.28	-0.027	0.30	0.29	-0.011
No sanitation in household	0.25	0.22	-0.029	0.21	0.23	0.024
Wealth index	0.25 (0.14)	0.25 (0.14)	0.008	0.25 (0.13)	0.24 (0.14)	-0.011
Government assistance	0.64	0.68	0.032	0.63	0.71	0.085
Attended at least 6 workshop sessions	0.84	0.86	0.019	0.82	0.80	-0.016
Mother completed middle school	0.28	0.3	0.028	0.31	0.32	0.004
Mother completed high school	0.27	0.29	0.026	0.29	0.24	-0.051
Mother works full time	0.41	0.39	-0.014	0.41	0.40	-0.011
Mother works part time	0.19	0.23	0.041	0.18	0.19	0.003
Father works full time	0.8	0.83	0.033	0.79	0.81	0.017
Any negative shock past 12 months	0.52	0.53	0.016	0.58	0.58	0.002
Discount rate	0.07 (0.08)	0.09 (0.09)	0.02	0.08 (0.09)	0.08 (0.09)	0.01
Parental Sense of Competence (efficacy)	4.13 (0.74)	4.18 (0.71)	0.05	4.12 (0.70)	4.10 (0.77)	-0.02
# negative shocks past 12 months	1.17 (1.30)	1.44 (1.61)	0.27	1.21 (1.36)	1.49 (1.60)	0.28
N	98	80		195	122	

Notes: Columns (1)-(3) refer to the sample with videos at baseline and follow-up and Columns (4)-(6) refer to the sample only with videos in follow-up. Column (1) and (4) show mean values for the variables in each row across subjects with ITT=0. Column (2) and (5) show mean values for the characteristics in each row for subjects with ITT=1. Standard deviations are reported in parentheses for non-dichotomous variables. Columns (3) and (6) show the differences in covariates at baseline by ITT status which are obtained by regressing the covariate on a dummy variable that takes the value of 1 if ITT=1 and of 0 if ITT=0. Due to missing values, the number of observations for each covariate may differ slightly from the N in the last row. Difference statistically significant at: ***1%, ** 5%, * 10%.

4.5 Descriptive statistics of language outcomes

Table 2 describes the main language outcomes we assess in the analysis. We present outcomes for control group families at follow-up, both for the sample with videos at baseline and follow-up, and for the sample with videos at follow-up, but not necessarily at baseline. On average, there are 112 child vocalizations in a 10-minute free play interaction between the child and the caregiver. This average is above the number of vocalizations found by Soderstrom and Wittebolle (2013) in a similar setting (they report 55 vocalizations on average). The total duration of vocalizations is 200 seconds in the first sample and 195 in the second one. Adults show around 110 vocalizations, with a total duration of 146 seconds in the first sample and 158 in the second one. These counts are similar to those found by Hurtado, Marchman and Fernald (2008) in a similar context of 10 minutes of organized playtime. There are around 90 conversational turns defined by a threshold of 0.3 seconds between vocalization off- and onset. Between 55 and 60 percent of the time, adults respond to children's vocalizations in less than 2 seconds. The average frequency of the adult pitch is around 5.5 Equivalent Rectangular Bandwidths (ERB), and its range (the maximum frequency minus the minimum one) is around 2.4 ERB.

Table 2: Description of language outcomes for control families at follow-up (N=98)

	Sample with videos at baseline & follow-up		Sample with videos at follow-up			
	Mean	Std. dev	Mean	Std.dev	Min	Max
	(1)	(2)	(3)	(4)	(5)	(6)
Count of child vocalizations	111.9	44.0	112.3	45.9	11	196
Duration of child vocalizations (seconds)	200.4	123.5	195.4	120.6	7	554
Count of adult vocalizations	108.2	51.6	111.0	51.7	16	236
Duration of adult vocalizations	145.6	99.0	158.4	105.9	10	400
Total vocalization duration	346.0	150.9	353.8	156.4	26	701
Turns (0.3 seconds)	88.9	53.1	92.8	54.5	2	206
Adult response ratio	0.6	0.2	0.6	0.2	0	1
Adult pitch average	5.5	0.8	5.6	0.7	3	7
Adult pitch range	2.4	0.8	2.5	0.8	1	5

We also assess correlations across language outcomes (see Appendix Table A1). Children's count and duration of vocalizations are strongly associated with each other ($r=.64$; $p<.01$) as with conversational turns ($r=.52$; $p<.01$ and $r=.46$; $p<.01$ respectively). Adult count and duration of vocalizations show also a strong positive bivariate correlation, and both are associated with conversational turns and response ratio. Interestingly, while the count of adult vocalizations is positively, although not too strongly associated with the count of child vocalizations ($r=.18$; $p<.01$), the duration of adult vocalizations is negatively associated with the duration of child vocalizations ($r=-.19$; $p<.05$), suggesting a possible crowding out effect given the proposed fixed time activity. Finally, adult pitch range is positively correlated with count ($r=.24$; $p<.01$) and duration ($r=.42$; $p<.01$) of adult vocalizations, but it is negatively correlated with the count of child vocalizations ($r=-.18$; $p<.05$). Adult pitch average is negatively correlated with the duration of child vocalizations ($r=-.37$; $p<.01$) and uncorrelated with adult vocalization count and duration.

4.6 Estimation approach

We run ordinary least squares (OLS) regressions of each language outcome on an indicator of having been randomly assigned to receive messages. The coefficient of the latter indicator provides us with the Intention to Treat (ITT) effect, or the effect of having been randomly assigned to receive messages, on

language outcomes. In our core analysis, we estimate ITT effects using the sample with videos both at baseline and follow-up. We then run robustness analyses on the sample with videos at follow-up (but not necessarily at baseline). While the first sample is smaller, it is more balanced across treatment and control subjects and allows us to estimate standard errors with more precision by accounting for the outcome at baseline. We also explore spillover effects by including in our estimations an indicator if the family was assigned to the control group but belonged to a center that was randomized into treatment. Note, however, that due to the small number of videos observed in this sub-sample, we do not expect estimates on spillovers to be precise, and do not emphasize this feature in the analysis.

To improve precision and address potential sample selection due to attrition, in our core specification we adjust for a set of covariates including: (i) the value of the outcome at baseline;¹³ (ii) the variable used to stratify the randomization at the CAIF center level (a dummy equal to one if the average level of education of mothers participating in *Crianza Positiva* at the center was at least middle school); (iii) covariates that were unbalanced after attrition at a significant level of 10% or less (the mother's age, and the number of other children in the household); and (iv) other covariates that could help improve the precision of the estimation, such as the child's age and gender, mother's education, whether the child lived with both biological parents, father's working status and the number of months elapsed between the date when the intervention began and the final assessment. For statistically significant outcomes, we report an adjusted p-value of the ITT coefficient after accounting for the clustering of observations around centers using wild bootstrapping (Cameron, Gelbach and Miller, 2008). We also adjust p-values for multiple hypothesis testing using the Romano Wolf algorithm (Clarke, Romano and Wolf 2020).

We check for robustness by adjusting for different combination of covariates in the main sample. In addition, we address the differential probability of being assigned to treatment and control arms by using inverse probability weighting: we estimate the probability that the family belongs to the treatment, conditional on the observed covariates at baseline, and then weight treated observations by the inverse of such probability, and control observations by the inverse of 1 minus such probability. We also run the analysis in the full sample with videos at follow-up, without adjusting for baseline outcomes.

¹³ To maximize sample size, we consider observations that had a recording in at least one of the two pre-intervention assessments (August 2017 or November-December 2017). We control for dummy variables that account for the timing of this assessment.

5. Results

5.1 Main specification

Our core specification estimates ITT effects for the sample with videos both at baseline and follow-up. Table 3 reports the effects of the intervention on the different language outcomes: measures of count and duration of vocalizations (in columns 1 to 5) and measures of conversational quality (in columns 6 to 9). At a 10% level of statistical significance, we find evidence that the intervention decreased the total duration of child vocalizations and increased the total duration of adult vocalizations. The estimates on the ITT indicator are -45 seconds per vocalization for the child, compared to an average duration for control children of 200 seconds (-23%), and 29 seconds for adults, compared to an average of 146 seconds for control subjects (20%). The finding that adults speak more is in line with our hypotheses, but not the finding that children vocalize less. We find no statistically significant effects over the summation of child and adult vocalizations (total vocalization duration), suggesting that parental vocalizations are crowding out child vocalizations. There is no evidence of effects on the count of child or adult vocalizations. Once we account for clustering of observations around centers, the effect on the duration of adult vocalizations loses statistical significance at traditional levels. And none of the effects are robust to adjustment for multiple hypotheses testing.

Regarding outcomes that assess communication quality, the first outcomes analyzed, conversational turns and parental responsiveness, show no differences across treated and control subjects. Coefficients appear to be imprecisely estimated. On the other hand, we find a strong and positive effect on a measure of speech quality: the range of adult pitch. This estimate is 0.4 higher for parents in the intervention group, 18% above the frequency range observed for control parents (2.4 ERB). The effect remains statistically significant at $p < .10$ after clustering standard errors around centers using wild bootstrapping techniques and at $p < .05$ after adjusting p-values for multiple hypotheses testing. We also find that the intervention had a positive and statistically significant effect at the 10% level on the range of adult pitch in the spillover sample, robust to wild bootstrapping but not to multiple hypotheses testing. These effects support our hypothesis that the intervention contributed to improve the quality of language input.

Table 3: ITT effects of CP-messages on language outcomes

	<u>Count and duration of vocalizations</u>					<u>Conversational quality</u>			
	Count of child vocalizations (1)	Duration of child vocalizations (seconds) (2)	Count of adult vocalizations (3)	Duration of adult vocalizations (seconds) (4)	Total vocalization duration (seconds) (5)	Number of turns (threshold of 0.3 seconds) (6)	Adult Response Ratio (7)	Adult Pitch Average (8)	Adult Pitch Range (9)
ITT coefficient	-10.266	-44.668*	9.218	28.993*	-16.555	-3.011	0.028	0.162	0.424***
ITT std. error	(8.060)	(22.819)	(8.261)	(17.265)	(27.502)	(9.451)	(0.038)	(0.133)	(0.147)
Wild bootstrapping p-value		[0.095]		[0.289]					[0.066]
Romano Wolf mht p-value		[0.193]		[0.269]					[0.016]
Spillover coefficient	-12.880	-35.642	25.764	50.041	18.079	6.863	0.095	0.035	0.405*
Spillover std. error	(12.015)	(38.433)	(15.655)	(33.307)	(42.005)	(15.320)	(0.064)	(0.214)	(0.235)
Wild bootstrapping p-value									[0.077]
Romano Wolf mht p-value									[0.308]
N	159	157	159	159	157	159	157	158	158
Outcome mean ITT=0	111.88	200.39	108.20	145.62	346.02	88.86	0.55	5.52	2.42
Outcome std. dev ITT=0	43.96	123.53	51.61	99.04	150.87	53.06	0.21	0.75	0.79

Notes: *p<0.1, **p<0.05, ***p<0.01. The table shows effects of randomization into treatment (ITT) or into the spillover sample on quantitative language outcomes after adjusting for the following set of covariates: randomization strata, outcome at baseline, mother's education, mother's age, child's gender, child's age in months when the intervention began, number of months elapsed between the final assessment and the date when the intervention began, number of people in the household, single child, whether the father worked full time at baseline, and indicators for the timing when the family was videotaped at baseline (in August 2017, in November-December 2017, or both). When there are two baseline measures, we use the average. Robust standard errors are reported in parentheses. For significant outcomes, we report in squared brackets the p-values after adjusting standard errors for clustering of observations around centers using wild bootstrapping. We also report p-values after adjusting for multiple hypothesis testing (mht) using the Romano Wolf algorithm and considering two families of outcomes: count and duration of vocalizations, and conversational quality.

5.2 Robustness

Table 4 shows ITT effects conditional on different sets of covariates. We include outcomes for which the effect was statistically significant in Table 3. Specification 1 replicates results in Table 3 (same coefficients as in columns (2),(4) and (9)). Specification 2 shows ITT effects after adjusting for sociodemographic status (SES) and the outcome variable at the last baseline survey, if available, or at the first baseline survey otherwise. The third specification shows ITT results adjusting for SES and the outcome at the first baseline survey, or at the second baseline survey otherwise. Specification 4 shows ITT effects when adjusting for the average of the outcome at the two baseline surveys only (no adjustment for SES). The final specification weights treatment and control observations for the inverse of the estimated propensity score of being in each group. We report wild bootstrapping p-values in square brackets. Results are very robust when assessing the effect of treatment on adult pitch range. Coefficients on the duration of child and adult vocalizations, on the other hand, are more sensitive to alternative specifications; and as in Table 3, first panel, most of the estimates lose statistical significance after adjusting standard errors for clustering.

Table 4: Robustness of ITT effects to different set of covariates

<i>Outcomes</i>	Duration of child vocalizations (1)	Duration of adult vocalizations (2)	Adult pitch Range (3)
<i>Alternative specifications</i>			
1) Covariates: average of SES and baseline outcome between baseline 0 and baseline 1 (N=159)			
ITT coefficient	-44.668*	28.993*	0.424***
ITT std. error	(22.819)	(17.265)	(0.147)
Wild bootstrapping p-value	[0.095]	[0.289]	[0.066]
2) Covariates: SES and baseline outcomes at baseline 1 (use baseline 0 when baseline 1 missing) (N=157)			
ITT coefficient	-44.079*	29.239*	0.408***
ITT std. error	(22.539)	(17.018)	(0.149)
Wild bootstrapping p-value	[0.119]	[0.280]	[0.080]
3) Covariates: SES and baseline outcomes at baseline 0 (use baseline 1 when baseline 0 missing) (N=157)			
ITT coefficient	-43.399*	27.863	0.428***
ITT std. error	(22.81)	(17.257)	(0.143)
Wild bootstrapping p-value	[1.261]		[0.070]
4) Covariates: average baseline outcome between baseline 0 and 1; no adjustment for SES (N=174)			
ITT coefficient	-26.124	30.392*	0.389***
ITT std. error	(19.767)	(16.06)	(0.133)
Wild bootstrapping p-value		[0.282]	[0.054]
5) Same as 1) + inverse probability weighting (N=145)			
ITT coefficient	-27.290	33.306*	0.412***
ITT std. error	(23.248)	(17.030)	(0.143)
Wild bootstrapping p-value		[0.240]	[0.054]

Notes: * p<0.1; **p<0.05; ***p<0.01. The table shows effects of randomization into treatment (ITT) on language outcomes after adjusting for different sets of covariates. Columns indicate different outcomes and rows specifications with different sets of covariates in the regressions. For each outcome/specification combination, we show the ITT coefficient from the regression, its robust standard error in parentheses, and in square brackets the p-values after clustering standard errors around centers using wild bootstrapping (WB p). Specification 1 shows ITT estimates after adjusting for sociodemographic characteristics (SES) and the average of the outcome measured at the two baseline surveys. Specification 2 shows ITT effects after adjusting for SES and the outcome variable at the last baseline survey if available or at the first baseline survey otherwise. Specification 3 shows ITT results adjusting for SES and the outcome at the first baseline survey or at the second one otherwise. Specification 4 shows ITT effects adjusting for the average of the outcome at the two baseline surveys, but without adjustment for SES. Specification 5 is similar to 1, but weights observation for the inverse probability of having been assigned to treatment.

In Appendix Table B1 we show results of estimating the Treatment Effect on the Treated (TOT), rather than Intention to Treat effects. We use Two Stages Least Square estimation techniques and instrument the probability of having received at least one message with the indicator of random assignment to treatment. Results are very similar to those in Table 3. Exposure to the messages decrease the duration of child vocalizations by 47 seconds (wild bootstrapping p-value = .09) and increases the duration of

adult vocalizations by 40 seconds, although this effect becomes marginally insignificant once we cluster standard errors around centers. Receiving parenting messages continues to show a strong positive effect on adult pitch range (0.50 ERB), which remains statistically significant at 5% after clustering standard errors.

Next, we estimate ITT results analyzing the video's last 300 seconds rather than the full video (which theoretically should have an extension of 600 seconds). We do so for two reasons: i) to measure outcomes in a similar timeframe for all observations, and ii) because we expect mother and child to be more relaxed and less influenced by socially desirable norms in the second half of the video. Results continue to show a positive effect of the intervention on the adult pitch range – of 0.43 ERB – and some evidence of an increase in the duration of adult vocalizations, which marginally loses statistical significance after adjusting standard errors for clustering (see Appendix Table B2).

In Appendix Table B3 we show results when running the ITT analysis for the full sample of families with a video recording at follow-up (regardless of having or not a video at baseline). Again, we identify a positive ITT effect on the adult pitch range, and a negative effect on the duration of child vocalizations, but the magnitudes are smaller than in the core specification and lose statistical significance after clustering standard errors around centers.

5.3 Mechanisms

We explore mechanisms in two ways. First, we investigate whether exposure to the message content led parents to increase their involvement in activities associated with language development. As mentioned before, 17% of messages encouraged parents to engage in language enhancing activities such as holding conversations with the child, asking open questions, describing objects, places and activities, reading to the child, or singing and listening to songs together. Furthermore, other messages were oriented towards strengthening parental-child interactions through the development of sensitivity and responsiveness, key features for good communication and language development.

In a post-treatment survey (administered between one and three months after the messaging intervention finished), we asked parents for their frequency of involvement in several activities with the child, including singing songs, reading, telling stories, and naming, telling, or drawing things. Each measure ranged from 1 to 7, where 1 was never in the week and 7 was every day or nearly every day. We standardized each outcome dividing it by the standard deviation of the outcome in the control group. In addition, we constructed an aggregate index of parental involvement in literacy activities using the first component of a principal component analysis on the six parental involvement outcomes. This component explained 53% of the overall variability in the outcomes and was also standardized. We regressed all standardized outcomes on the ITT indicator, adjusting for the same set of covariates as in

our core specification¹⁴. Results are reported in Table 5. Being assigned to treatment increased the frequency of reading books, telling stories, and naming, telling or drawing things to the child by 0.30, 0.27 and 0.44 standard deviations respectively. The overall effect on the aggregate index is of 0.37 standard deviations. All effect are robust to wild bootstrapping.

Table 5: Effects of CP-messages on parental frequency of involvement in literacy activities

	I sing songs to my child	My child and I sing songs together	I read books to my child or we look at images	I tell stories to my child	I name, tell or draw things to my child	Index of parental involvement in literacy activities
	(1)	(2)	(3)	(4)	(5)	(6)
ITT standardized coefficient (d)	0.117	0.172	0.295**	0.271*	0.437***	0.374***
ITT std. error	(0.152)	(0.158)	(0.143)	(0.147)	(0.158)	(0.136)
Wild bootstrapping p-value			[0.077]	[0.047]	[0.007]	[0.012]
N	145	145	145	141	152	133

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Each outcome is standardized (expressed in terms of the standard deviation of the outcome in the control sample). Robust standard errors in parentheses and p-value after clustering standard errors with wild bootstrapping in squared brackets (only for significant outcomes). Regressions adjust for strata, an indicator of non-treated families within treated centers, outcome at baseline, mother's education, mother's age, child's gender, child's age in months when the intervention began, number of months elapsed between the final assessment and the date when the intervention began, number of people in the household, single child, and whether the father worked full time at baseline. Sample includes observations with videos at baseline and follow-up.

We also investigate whether these parental involvement outcomes correlate with the pitch range and the duration of adult vocalization at follow-up, the two parental language outcomes for which we find a positive effect of the intervention. We find a correlation of $r = .13$ ($p < .1$) between telling stories to the child and the adult pitch range. We do not find a statistically significant association between the measures of parental engagement and adult vocalization duration, but do find a significant correlation between the aggregate index of parental engagement and the total duration of vocalizations ($r = .14$; $p < .1$).

Next, we explore whether ITT effects are heterogeneous across parents with different degrees of behavioral biases. Recall that the messages aimed at addressing behavioral barriers such as present bias, cognitive fatigue, and negative identities. The first message of the week presented information and reminded parents on the importance of a specific parenting practice (making future benefits more

¹⁴ This includes strata indicators, an indicator of being part of the spillover sample, outcomes at baseline and a set of covariates capturing family socioeconomic status.

salient); the second one addressed cognitive fatigue and inattention by proposing simple and concrete activities to do during the week; and the third one targeted parental identities, via messages of encouragement and reflection. If the messages work by circumventing these behavioral biases, we would expect parents with stronger biases to react more to the treatment. As shown in Table 1, we measured these dimensions by using parents' discount rate (present bias), negative shocks faced by the family (cognitive fatigue), and parents' sense of competence (parental identity). We expect the treatment to have stronger impacts among parents with higher discount rates, households with more negative shocks, and parents' with lower sense of competence. For each language outcome that was affected by the intervention, we run a regression that adds to the core specification a main effect for a behavioral barrier and an interaction between this barrier and the ITT variable (see Table 6). Rather than using the measures for discount rate, negative shocks, and parental efficacy displayed in Table 1, we work with dichotomous indicators of these variables being above (discount rate, negative shocks) or below (parental efficacy) the mean. We find that parents with high discount rates or low efficacy are more likely to show increases in the pitch range after treatment at a $p < .05$ and $p < .1$ respectively. Both effects are robust to wild-bootstrapped standard errors and have the expected sign. Impatient parents are more likely to react to messages that remind them about the benefits of speaking and reading to the child, and of responding sensibly to the child's cues. Parents with low sense of efficacy are more likely to react to messages of encouragement and support. The coefficient on the other interaction (ITT times parental cognitive fatigue) is also positive, but statistically insignificant.

Table 6: Heterogeneity of ITT effects by behavioral barrier

	Duration of child vocalizations (1)	Duration of adult vocalizations (2)	Adult pitch range (3)
<i>Present Bias</i>			
ITT	-59.197*	17.052	0.170
	(31.302)	(22.734)	(0.164)
High discount rate	20.612	-14.791	-0.065
	(32.754)	(25.298)	(0.209)
ITT*High discount rate	9.249	33.968	0.583**
	(46.708)	(36.437)	(0.267)
			[0.080]
N	147	149	148
<i>Cognitive fatigue</i>			
ITT	-35.804	8.938	0.354**
	(28.414)	(20.531)	(0.162)
Negative shocks	19.197	12.702	0.200
	(28.546)	(22.526)	(0.201)
ITT*Negative shocks	-21.237	44.682	0.198
	(39.641)	(34.462)	(0.276)
N	155	157	156
<i>Negative identities</i>			
ITT	-52.683*	36.753	0.144
	(31.292)	(24.875)	(0.186)
Low parental efficacy	-2.959	-4.829	-0.273
	(30.961)	(24.468)	(0.188)
ITT*Low parental efficacy	54.186	-26.325	0.474*
	(44.706)	(36.614)	(0.245)
			[0.058]
N	140	142	141

Notes: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Robust standard errors in parentheses. When the interaction of interest is statistically significant, we report in squared brackets p-values after adjusting for the clustering of observations around centers using wild bootstrapping. The estimation sample includes observations that have a video in baseline and follow-up. Each pair of rows denotes a specific model in which the ITT is interacted with a behavioral barrier measured at baseline. The first two rows correspond to high present bias, proxied with a discount rate above the mean. The next couple of rows correspond to cognitive fatigue. The variable that measures the latter behavioral barrier is an indicator that takes value of 1 if the family was exposed to 2 or more negative shocks in the previous 12 months to the survey. The last two rows correspond to negative identities. The variable that measures the latter behavioral barrier is an indicator that takes the value 1 if the parental efficacy at baseline is less or equal than the mean. Each column denotes a different outcome. For each outcome, we show the coefficient and standard error for the ITT main effect, and the coefficient and standard error for the interaction between ITT and the behavioral barrier analyzed. The estimations control for the main effect of the behavioral barrier analyzed, an indicator of whether the subject was randomized to control within treated CAIF centers (the “spillover” sample), for randomization strata, SES, and outcomes at baseline.

6. Discussion and Conclusions

In this paper, we assess the effects of the *Crianza Positiva* text and audio e-messaging program on caregiver-child language interaction patterns. The *Crianza Positiva* e-messaging program is a 3-times-a-week, six-month-long intervention for families with children aged 0-2 years aimed at improving parenting competences. Its design recognizes that parental biases, such as present bias, inattention and cognitive fatigue, and negative identities, may lead to sub-optimal parenting practices, and exploits behavioral tools such as reminders, suggestions of actions, and messages of encouragement to reinforce and sustain positive parenting practices.

Crianza Positiva messages address the parental-child interaction by reinforcing sensitive observation, response, caring, and stimulation. These features are key components of a high quality communication between the child and the caregiver, and set the grounds for language development. In addition, several modules of the *Crianza Positiva* e-messaging program (almost one out of five of the messages) are aimed specifically at language stimulation, encouraging parents to engage in conversations with the child, ask open questions, describe objects, places and activities, read books, tell stories and sing to and with the child.

We evaluate the program using a randomized controlled trial in 24 early childhood centers in Uruguay. Families that had completed an 8-week workshop at the center were randomized into receiving or not receiving mobile messages. After the program, we videotaped 10-minutes-sessions of free play between the caregiver and the child, and decoded language patterns using automated techniques. We find that the intervention was successful at influencing parental behavior: it increased both the duration and the quality of parental vocalizations, as captured by pitch range. While the effect on the duration of parental vocalization was sensitive to alternative specifications, the effect on pitch range was very robust. Higher pitch has been associated with better infant-adult connection both in individual variation (e.g., comparison of depressed and non-depressed mothers; Porritt et al., 2014) and in experimental settings (i.e., when the mother is shown a supposedly-live video of her infant, but the infant can versus cannot actually hear the mother; Smith & Trainor, 2008). Furthermore, studies show that speech conveying pitch variations (larger ranges of speaker intonations) also results in higher levels of attention, more word recall and more recognition of information by the listener than those with homogenous pitch ranges across the sentences (Rodero, Potter and Prieto, 2017). *Parentese* speech, or speech characterized by higher pitch, slower tempo, and exaggerated intonation ranges, has been positively associated with early spoken word recognition and better acquisition of language by infants, and has been shown to be preferred by children relative to standard speech (Singh et al., 2009; Liu, Tsao & Kuhl., 2009; Ramírez-Esparza, García-Sierra & Kuhl, 2017). In turn, better language outcomes are related to better long run educational achievement and earnings (Gertler et al, 2014). These results are

promising because pitch range is not a variable parents can easily and consciously manipulate, and thus it could be less sensitive to desirability biases.

Consistent with parental language results, we find that the intervention also increased parental involvement in reading, telling stories and describing things to the child, as self-reported by parents in a follow-up survey. Furthermore, we find a positive correlation between telling stories and parental pitch range, as well as between an aggregate index of parental engagement in literacy activities and the total duration of vocalizations. The findings above suggest that the effects on *Parentese* speech and adult vocalization duration may have been mediated by an impact of the intervention on parental engagement in language interactions with the child and language stimulating activities. We also explore heterogeneous effects of the intervention by parental characteristics that may be capturing behavioral biases. We find evidence that the messages were most effective on parents with high discount rates, supporting our hypothesis that reminders help present-oriented parents to make more salient the benefits of positive parental engagement with the child.

It is important to bear in mind that our study may be among the first to analyze short video-recordings automatically in the context of large-scale studies (see Background, Sections 2.1-2). We therefore first checked our proposed metrics in terms of their reliability against manual coding, as well as external validity (i.e., correlations with age, which is our best proxy for language development in this sample; see Section 4.2). Our results suggested low external validity of child measures, as only one of our measures correlated significantly with child age (a proxy for language development available in the same data set). The one measure that had some external validity was child vocalization duration, but this measure exhibited a decrease among the treated children as compared to those in the control. Since we observe that there is basically a trade-off between adult speech duration and child speech duration, we attribute this negative result to a crowding out effect by the caregiver in the context of a fixed 10-minute suggested activity and a more pro-active parental role. We believe the most reasonable interpretation for this effect is not a negative effect of the intervention on the child, but rather an unfortunate masking of potential effects on the child due to greater caregiver engagement. This suggests that our analyses, while sufficient to reveal changes in parental behavior, may not be ideal to measure potential positive effects on the child. Instead, perhaps qualitative analyses of children's vocalization content (such as estimation of syntactic complexity and lexical diversity) may be more sensitive. Unfortunately, given the state of automatic speech recognition, it is not feasible to automatically generate transcriptions of these videos. Transcriptions for the 59 hours of data¹⁵ in the smaller sample would have required at least 1770 hours of human work¹⁶, which was also not feasible.

¹⁵ 178 dyads x 2 waves x 10 minutes = 3560 minutes or 59 hours.

¹⁶ 59 x 30 = 1770. The transcription of 1 minute of adult vocalization content requires approximately 20 minutes, while transcription of 1 minute child's vocalization content involves approximately 40 minutes.

Our work contributes to a nascent literature exploiting text-based communication and behavioral economic insights to boost early childhood development. It is worth emphasizing that in our intervention, messages were sent after completing a workshop with the families, so compliance and message efficacy may have been larger due to this prior contact. Moreover, we also add value to the literature by rigorously measuring language outcomes, which are seldom measured in the field due to the methodological difficulties with caregiver reporting and child observation (Weber et al., 2017). Here, via the use of videos and automatic transcriptions, we are able to identify some interesting outcomes previously unexplored in the literature, such as parental intonation or pitch range. Our results demonstrate the large potential of these very low cost interventions, based on mobile technology and on the understanding of behavioral biases, in changing and sustaining parental behavioral patterns that ultimately lead to improved child development.

References

- Aboud, F. E., D. R. Singla, M. I. Nahil, and I. Borisova. 2013. "Effectiveness of a Parenting Program in Bangladesh to Address Early Childhood Health, Growth and Development." *Social Science and Medicine* 97(November): 250–58.
- Allen, S. E., & Dench, C. (2015). Calculating mean length of utterance for Eastern Canadian Inuktitut. *First Language*, 35(4-5), 377-406.
- Álvarez, M., Byrne, S., & Rodrigo, M. J. (2021). Social support dimensions predict parental outcomes in a Spanish early intervention program for positive parenting. *Children and Youth Services Review*, 121, 105823
- Bloomfield, J., Balsa, A., Cid, A. (2020). Using Behavioral Insights in Early Childhood Interventions: the Effects of Crianza Positiva E-Messaging Program on Parental Investment. *Working Paper*. Universidad de Montevideo.
- Ajzenman, N., & López Bóo, F. (2019). Lessons from Behavioral Economics to Improve Treatment Adherence in Parenting Programs: An Application to SMS. *IZA Discussion Paper No. 12808* <https://www.iza.org/publications/dp/12808/lessons-from-behavioral-economics-to-improve-treatment-adherence-in-parenting-programs-an-application-to-sms>
- Baird, S., Bohren, J.A., McIntosh, C., & Özler, B. (2018). Optimal design of experiments in the presence of interference. *Review of Economics and Statistics* 100(5), 844–860.
- Bayley, N. (2006). *Bayley Scales of Infant and Toddler Development: Administration Manual*. 3rd. United States of America: Psychorp.
- Bénabou, R., & Tirole, J. (2002). Self-confidence and personal motivation. *The Quarterly Journal of Economics* 117(3), 871–915.
- Berlinski, S., & Schady, N. (2015). More bang for the buck: investing in early childhood development. In *The Early Years* (pp. 149-178). Palgrave Macmillan, New York.
- Bornstein, M. H., Putnick, D. L., Cote, L. R., Haynes, O. M. & Suwalsky, J. T. D. (2015). Mother-Infant Contingent Vocalizations in 11 Countries. *Psychological Science*, 26(8), 1272-1284.
- Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3), 414-427.
- Clarke, D., Romano, J. P., & Wolf, M. (2020). The Romano–Wolf multiple-hypothesis correction in Stata. *The Stata Journal*, 20(4), 812-843.
- Conti, G., Poupakis, S., Sandner, M. & Kliem, S. (2020). The effects of home visiting on mother-child interactions: evidence from a randomized trial using dynamic micro-level data. *IZA DP No. 12937*
- Cristia, A. (2013). Input to language: The phonetics and perception of infant-directed speech. *Language and Linguistics Compass*, 7, 157-170. Doss, C., Fahle, E. M., Loeb, S., & York, B. N. (2019). More Than Just a Nudge Supporting Kindergarten Parents with Differentiated and Personalized Text Messages. *Journal of Human Resources*, 54(3), 567-603.
- Dale, A., & Strauss, A. (2009). Don't forget to vote: Text message reminders as a mobilization tool. *American Journal of Political Science*, 53(4), 787-804.
- Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., Pethick, S. & Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual*. San Diego: Singular Publishing Group.

- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J. & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59(5), 1-173.
- Ferjan Ramírez, N., Lytle, S. R., Fish, M., & Kuhl, P. K. (2019). Parent coaching at 6 and 10 months improves language outcomes at 14 months: A randomized controlled trial. *Developmental Science*, 22(3), e12762.
- Ferjan Ramírez, N., Lytle, S. R., & Kuhl, P. K. (2020). Parent coaching increases conversational turns and advances infant language development. *Proceedings of the National Academy of Sciences*, 117(7), 3484-3491.
- Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, 10(3), 279-293.
- Fernald, L., Kariger, P., Engle, P. & Raikes, A. (2009). *Examining child development in low-income countries: a toolkit for the assessment of children in the first five years of life*. The World Bank, Washington, DC.
- Fernald, A., Marchman, V. A. & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, 16(2), 234-248.
- Frankenburg, W. K., Dodds, J., Archer, P., Bresnick, B., Maschka, P., Edelmann, N. & Shapiro, H. (1990). *The Denver II technical manual Developmental Materials*.
- Gennetian, L., Darling, M., and Aber, J. L. 2016. "Behavioral Economics and Developmental Science: A New Framework to Support Early Childhood Interventions." *Journal of Applied Research on Children*, 7(2): 2.
- Gertler, P., Heckman, J., Pinto, R., Zanolini, A., Vermeerch, C., Walker, S, Chang, S.M. & Grantham-McGregor, S. (2014). Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica. *Science*, 344(6187), 998-1001.
- Gilkerson, J., Richards, J. A., Warren, S. F., Montgomery, J. K., Greenwood, C. R., Kimbrough Oller, D., ... & Paul, T. D. (2017). Mapping the early language environment using all-day recordings and automated analysis. *American Journal of Speech-language Pathology*, 26(2), 248-265.
- Gurol-Urganci, I., de Jongh, T., Vodopivec-Jamsek, V., Atun, R., & Car, J. (2013). Mobile phone messaging reminders for attendance at healthcare appointments. *Cochrane Database of Systematic Reviews*, (12).
- Hart, B. & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H Brookers Publishing.
- Hurtado, N., Marchman, V. A. & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science*, 11(6), 31-39.
- Hurwitz, L. B., Lauricella, A. R., Hanson, A., Raden, A., & Wartella, E. (2015). Supporting Head Start parents: impact of a text message intervention on parent-child activity engagement. *Early Child Development and Care*, 185(9), 1373-1389.
- Johnston, Charlotte, and Eric J. Mash. 1989. "A measure of parenting satisfaction and efficacy." *Journal of Clinical Child Psychology* 18(2):167-175.
- Karp, S. M., Lutenbacher, M., & Wallston, K. A. (2015). Evaluation of the parenting sense of competence scale in mothers of infants. *Journal of Child and Family Studies*, 24(11), 3474-3481.

- Kirby, Kris N., Nancy M. Petry, and Warren K. Bickel. 1999. "Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls." *Journal of Experimental Psychology: General* 128(1):78–87.
- Lavechin, M., de Seyssel, M., Gautheron, L., Dupoux, E., & Cristia, A. (submitted). Reverse-engineering language acquisition with child-centered long-form recordings. *Annual Review of Linguistics*.
- Law, J., Rush, R., Schoon, I. & Parsons, S. (2009). Modeling developmental language difficulties from school entry into adulthood: literacy, mental health, and employment outcomes. *The Journal of Speech, Language, and Hearing Research*, 52(6), 1401-16.
- Le Franc, A., Riebling, E., Karadayi, J., Wang, Y., Scaff, C., Metze, F., & Cristia, A. (2018). The ACLEW DiViMe: An Easy-to-use Diarization Tool. *Interspeech*, 1383-87.
- Leung, C. Y. Y., Hernandez, M. W. & Suskind, D. L. (2020). Enriching home language environment among families from low-SES backgrounds: A randomized controlled trial of a home visiting curriculum. *Early Childhood Research Quarterly*, 50, 24-35.
- Liu, H. M., Tsao, F. M., and Kuhl, P. K. (2009). Age related changes in acoustic modifications of Mandarin maternal speech to preverbal infants and five-year-old children: a longitudinal study. *Journal of Child Language*: 36, 909–922.
- López Bóo, F., Cubides Mateus, M., Sorio, R., Garibotto, G., & Berón, C. (2018). Measuring the Quality of the Home Environment of Young Children in Uruguay: Socioeconomic Gradients in the Home Inventory. *Inter American Development Bank Technical Notes*. IDB-TN-1550.
- Love, J. M., Kisker, E. E., Ross, C., Raikes, H., Constantine, J., Boller, K., ... & Fuligni, A. S. (2005). The effectiveness of early head start for 3-year-old children and their parents: lessons for policy and programs. *Developmental Psychology*, 41(6), 885.
- Mani, A., Mullainathan, S., Shafir, E. and Zhao, J. 2013. Poverty impedes cognitive function. *Science*, 341(6149): 976–980.
- Mayer, Susan E., Ariel Kalil, Philip Oreopoulos, and Sebastián Gallegos. (2018). Using behavioral insights to increase parental engagement: The parents and children together intervention. *Journal of Human Resources* 0617–8835R.
- Mendive, S., Mascareño Lara, M., Aldoney, D., Pérez, J. C., & Pezoa, J. P. (2020). Home Language and Literacy Environments and Early Literacy Trajectories of Low-Socioeconomic Status Chilean Children. *Child Development* 91(6), 2042-2062.
- McGillion, M., Pine, J. M., Herbert, J. S., Matthews, D. (2017). A randomized controlled trial to test the effect of promoting caregiver contingent talk on language development in infants from diverse socioeconomic status backgrounds. *Journal of Child Psychology and Psychiatry*, 58(10), 1122-1131.
- Meuwissen, Alyssa, Alison Giovanelli, Madelyn Labella, and Amy Susman-Stillman. (2017). Text2Learn: An Early Literacy Texting Intervention by Community Organizations.
- Neumann, Christof (2020). Avutils. Available from <https://github.com/gobbios/avutils>.
- NICHD. (2000). The relation of child care to cognitive and language development. *Child Development*, 71(4), 960-980
- Olds, D. L., Kitzman, H., Cole, R., Robinson, J., Sidora, K., Luckey, D. W., ... & Holmberg, J. (2004). Effects of nurse home-visiting on maternal life course and child development: age 6 follow-up results of a randomized trial. *Pediatrics*, 114(6), 1550-1559.

- Phillips, C. B., Hurley, J. C., Angadi, S. S., Todd, M., Berardi, V., Hovell, M. F., & Adams, M. A. (2020). Delay discount rate moderates a physical activity intervention testing immediate rewards. *Behavioral Medicine, 46*(2), 142-152.
- Porritt, L. L., Zinser, M. C., Bachorowski, J. A., & Kaplan, P. S. (2014). Depression diagnoses and fundamental frequency-based acoustic cues in maternal infant-directed speech. *Language Learning and Development, 10*(1), 51-67.
- Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2017). Look who's talking NOW! Parentese speech, social context, and language development across time. *Frontiers in Psychology, 8*, 1008.
- Räsänen, O., Seshadri, S., Karadayi, J., Riebling, E., Bunce, J., Cristia, A., ... & Soderstrom, M. (2019). Automatic word count estimation from daylong child-centered recordings in various language environments using language-independent syllabification of speech. *Speech Communication, 113*, 63-80.
- Rodero, E., Potter, R. F., & Prieto, P. (2017). Pitch range variations improve cognitive processing of audio messages. *Human Communication Research, 43*(3), 397-413.
- Rodriguez, M. L., Dumont, K., Mitchell-Herzfeld, S. D., Walden, N. J., & Greene, R. (2010). Effects of Healthy Families New York on the promotion of maternal parenting competencies and the prevention of harsh parenting. *Child Abuse & Neglect, 34*(10), 711-723.
- Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M.L., & Gabrieli, J. D. E. (2018). Beyond the 30-million-word gap: children's conversational exposure is associated with language-related brain function. *Psychological Science, 29*(5), 700-710.
- Rowe, M. L., Raudenbush, S. W. & Goldin-Meadow, S. (2012). The pace of vocabulary growth helps predict later vocabulary skill. *Child Development, 83*(2), 508-25.
- Rubio-Codina, M., & Grantham-McGregor, S. (2020). Validez predictiva de pruebas cortas comúnmente usadas para medir el desarrollo infantil en estudios a gran escala. *Inter-American Development Bank Working Paper*. DB-WP-1174.
- Schachner, A. & Hannon, E. E. (2011). Infant-directed speech drives social preferences in 5-month-old infants. *Developmental Psychology, 47*(1), 19-25.
- Schilbach, F., Schofield, H. and Mullainathan, S. (2016). The psychological lives of the poor. *American Economic Review, 106*(5), 435-40.
- Shah, A. K., Mullainathan, S., and Shafir, E. 2012. Some consequences of having too little. *Science 338*(6107), 682-685.
- Shneidman, L. A., & Goldin-Meadow, S. (2012). Language input and acquisition in a Mayan village: How important is directed speech?. *Developmental Science, 15*(5), 659-673.
- Singh, L., Nestor, S., Parikh, C., and Yull, A. (2009). Influences of infant-directed speech in early word recognition. *Infancy, 14*, 654-666.
- Smith, N. A., & Trainor, L. J. (2008). Infant-directed speech is modulated by infant feedback. *Infancy, 13*(4), 410-420.
- Snow, C. E. & Ferguson, C. A. (eds). (1977). *Talking to children*. Cambridge University Press.
- Soderstrom, M., & Wittebolle, K. (2013). When do caregivers talk? The influences of activity and time of day on caregiver speech and child vocalizations in two childcare environments. *PloS one, 8*(11), e80646.
- Sperry, D. E., Sperry, L. L. & Miller, P. J. (2019). Reexamining the Verbal Environments of Children From Different Socioeconomic Backgrounds. *Child Development, 90*(4), 1303-1318.

- Thaler, Richard H. (2015). *Misbehaving: The making of behavioral economics*. New York: WW Norton.
- The ManyBabies Consortium (2020). *Quantifying Sources of Variability in Infancy Research Using the Infant-Directed-Speech Preference*. *Advances in Methods and Practices in Psychological Science*.
- Tomasello, M. & Stahl, D. (2004). Sampling children's spontaneous speech: how much is enough? *Journal of Child Language*, 31(1), 101-121.
- Weber, A., Fernald, A., & Diop, Y. (2017). When cultural norms discourage talking to babies: Effectiveness of a parenting program in rural Senegal. *Child Development*, 88(5), 1513-1526.
- Weisleder, A. & Fernald, A. (2013). Talking to children matters: early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24(11), 2143-2152.
- Zangl, R. & Mills, D. L. (2007). Increased Brain Activity to Infant-Directed in 6-and 13-Month-Old Infants. *Infancy*, 11(1), 31-62.
- Zimmerman, F. J., Gilkerson, J., Richards, J. A., Christakis, D. A., Xu, D., Gray, S. & Yapanel, U. (2009). Teaching by listening: the importance of adult-child conversations to language development. *Pediatrics*, 124(1), 342-349.

Appendix A: Correlations

Figure A: Pearson Correlations between Manual and Automatic Annotations. Main Outcomes.

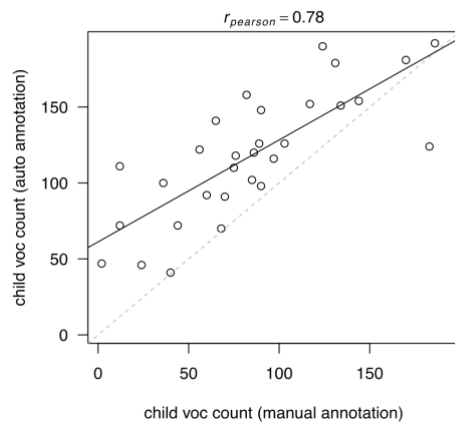


Figure A1: Count of child vocalizations

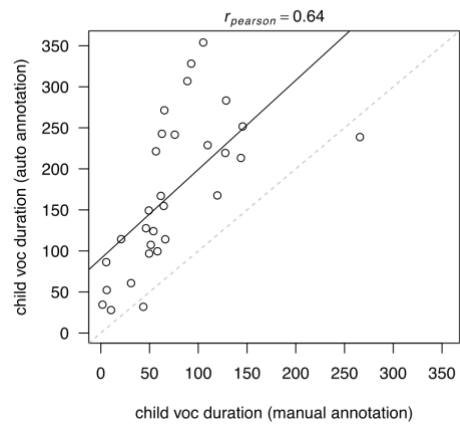


Figure A2: Duration of child vocalizations

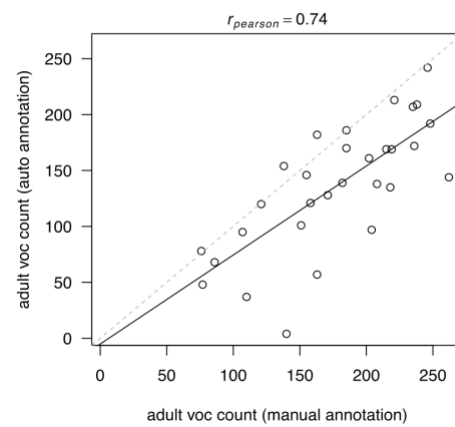


Figure A3: Count of adult vocalizations

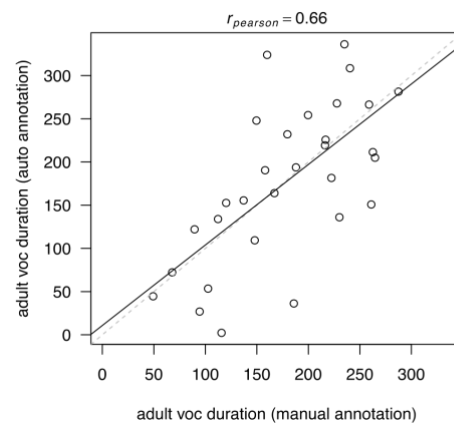


Figure A4: Duration of adult vocalizations

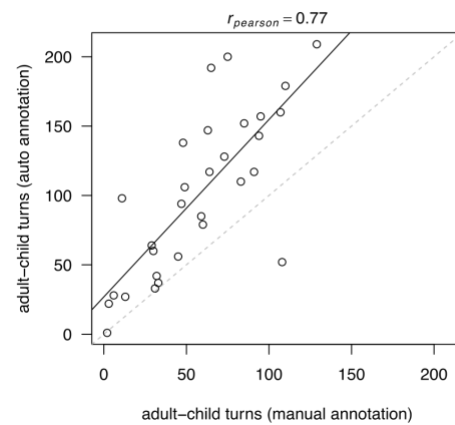


Figure A5: Adult-child turns

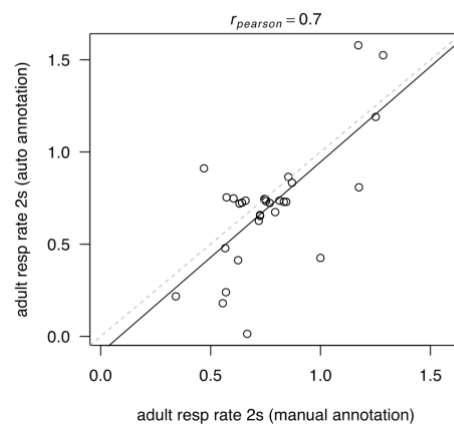


Figure A6: Adult response rate 2 second

Table A1: Pairwise correlations across outcomes

	Count of child vocalizations	Duration of child vocalizations (seconds)	Count of adult vocalizations	Duration of adult vocalizations	Total vocalization duration	Turns (0.3 seconds)	Adult response ratio	Adult pitch average	Adult pitch range
Count of child vocalizations	1								
Duration of child vocalizations (seconds)	0.638***	1							
Count of adult vocalizations	0.177**	-0.123	1						
Duration of adult vocalizations	-0.044	-0.192**	0.840***	1					
Total vocalization duration	0.514***	0.718***	0.491***	0.546***	1				
Turns (0.3 seconds)	0.524***	0.457***	0.659***	0.594***	0.811***	1			
Adult response ratio	-0.0625	-0.0367	0.805***	0.838***	0.563***	0.709***	1		
Adult pitch average	-0.0355	-0.367***	0.022	-0.019	-0.326***	-0.158**	-0.073	1	
Adult pitch range	-0.181**	-0.0889	0.241***	0.422***	0.224***	0.238***	0.396***	0.031	1

Appendix B: Robustness estimations

Appendix Table B1: Treatment on the treated (TOT) effects of CP-messages on language outcomes

	Count of child vocaliza- tions	Duration of child vocaliza- tions (seconds)	Count of adult vocali- zations	Duration of adult vocali- zations (seconds)	Total vocaliz- ation duration (seconds)	Number of turns (thres- hold of 0.3 seconds)	Adult Response Ratio	Adult Pitch Average	Adult Pitch Range
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
TOT coefficient	-12.675	-46.451**	14.170*	39.898**	-6.709	4.954	0.064*	0.134	0.504***
TOT std. error	(7.851)	(22.049)	(8.372)	(17.478)	(27.358)	(9.429)	(0.037)	(0.129)	(0.147)
Wild bootstrapping p-value		[0.092]	[0.394]	[0.141]			[0.302]		[0.021]
N	159	157	159	159	157	159	157	158	158

Appendix Table B2: Intention to treat (ITT) effects of CP-messages on language outcomes in last 300 seconds of video

	Count of child vocalizations	Duration of child vocalizations (seconds)	Count of adult vocalizations	Duration of adult vocalizations (seconds)	Total vocalization duration (seconds)	Number of turns (threshold of 0.3 seconds)	Adult Response Ratio	Adult Pitch Average	Adult Pitch Range
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ITT coefficient	-3.914	-18.679	7.099	21.912*	14.586	-0.628	0.032	0.158	0.428**
ITT std. error	(5.563)	(12.304)	(7.170)	(12.473)	(16.770)	(6.864)	(0.058)	(0.165)	(0.191)
Wild bootstrapping p-value				[0.114]					[0.063]
N	159	157	159	159	157	159	157	158	158

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The table shows Intention to Treat (ITT) effects on language outcomes measured in last 300 seconds of video after adjusting for the following covariates: randomization strata, outcome at baseline, mother's education, mother's age, child's gender, child's age in months when the intervention began, number of months elapsed between the final assessment and the date when the intervention began, number of people in the household, single child, whether the father worked full time at baseline, and indicators for the timing when the family was videotaped at baseline (in August 2017, in November-December 2017, or both). When there are two baseline measures, we use the average. Robust standard errors are reported in parentheses. For significant outcomes, we report in squared brackets the p-value on the ITT coefficient after adjusting standard errors for clustering around centers using wild bootstrapping.

Appendix Table B3: ITT effects on larger sample with videos at follow-up

<i>Outcomes</i>	Duration of child vocalizations	Duration of adult vocalizations	Adult pitch Range
	(1)	(2)	(3)
1) Adjusting for SES (N=274)			
ITT coefficient	-28.968*	8.641	0.299***
ITT std. error	(16.211)	(13.661)	(0.114)
Wild bootstrapping p-value	[0.298]		[0.143]
2) Adjusting for SES + inverse probability weighting (N=247)			
ITT coefficient	-41.776**	4.042	0.257**
ITT std. error	(17.028)	(14.743)	(0.125)
Wild bootstrapping p-value	[p=0.173]		[p=0.201]

Notes: *p<0.1, **p<0.05, ***p<0.01. The table shows effects of randomization into treatment (ITT) on language outcomes using a larger sample with videos at follow-up but not necessarily at baseline. Regressions adjust for the following set of covariates: randomization strata, mother's education, mother's age, child's gender, child's age in months when the intervention began, number of months elapsed between the final assessment and the date when the intervention began, number of people in the household, single child, and whether the father worked full time at baseline. Robust standard errors are reported in parentheses. For significant outcomes, we report in squared brackets the p-values after adjusting standard errors for clustering of observations around centers using wild bootstrapping (WB p).