NBER WORKING PAPER SERIES

IT ALL STARTS WITH BELIEFS:
ADDRESSING THE ROOTS OF EDUCATIONAL INEQUITIES
BY SHIFTING PARENTAL BELIEFS

John A. List
Julie Pernaudet
Dana Suskind

Working Paper 29394
http://www.nber.org/papers/w29394

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2021

The different studies used in this paper have been funded by the W. K. Kellogg Foundation, the Pritzker Traubert Family Foundation, the PNC Foundation, the Heising-Simons Foundation, and the Hymen Milgrom Supporting Organization. It received approval from the University of Chicago Institutional Review Board. We are grateful to Amanda Bezik, Alison Hundertmark, Kristin Leffel, Christy Leung, and Michelle Saenz for the implementation and monitoring of the projects. We also thank participants of the ASSA 2021 meeting, the ESA 2019 conference, the AFE 2019 conference and the RCT Days 2019 conference for their useful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

It All Starts with Beliefs: Addressing the Roots of Educational Inequities by Shifting Parental Beliefs
John A. List, Julie Pernaudet, and Dana Suskind
NBER Working Paper No. 29394
October 2021
JEL No. C93,D83,I21,J13

ABSTRACT

Socioeconomic inequalities in child development crystallize at early stages, with associated disparities in parental investment in children. A key to understanding the data patterns is to document the sources underlying the observed inequalities. We first show that there are dramatic differences in parental beliefs across socioeconomic backgrounds (SES), with parents of higher SES being more likely to believe that parental investments impact child development. We then use two field experiments targeted to low-SES families to explore the mutability of such beliefs and their link to parental investments. In both cases, we find that parental beliefs about child development are malleable. The less intensive version of the program based on educational videos changes parental beliefs, but fails to lastingly increase parental investments and child outcomes. By contrast, in the more intensive version of our program combining home visits and feedback, the augmented beliefs are associated with enriched parent-child interactions and improved vocabulary, math, and social-emotional skills for the children. Together, these results suggest that changing parental beliefs can be an important pathway to raising parental investments and reducing socioeconomic gaps in children's skills, but that simple informational policies may not be sufficient.

John A. List
Department of Economics
University of Chicago
1126 East 59th
Chicago, IL 60637
and NBER
jlist@uchicago.edu

Julie Pernaudet
Department of Economics
University of Chicago
1126 East 59th
Chicago, IL 60637
julie.pernaudet@gmail.com

Dana Suskind
Department of Surgery
University of Chicago Medicine
5841 South Maryland Avenue
MC 1035
Chicago, IL 60637
dsuskind@surgery.bsd.uchicago.edu

1 Introduction

Economic growth over the past 1000 years can be viewed as sporadic, but a persistent trend is that over the past few centuries, countries of the Organisation for Economic Co-operation and Development (OECD) have fared better than their peers (Boltho and Toniolo (1999)). For centuries, scholars have explored theories to explain the causes of economic growth, with investment in human capital (Lucas (1988)) representing a key tenet more recently. Interestingly, the last several decades have also witnessed a consensus of thought on human capital investment as a key driver of income growth and inequality at the individual level (Becker (1993)). Indeed, within certain circles, scholars have argued that the most important investment society can make in its citizenry is to increase dramatically their investments in early childhood education (Heckman (2012)).

What remains unsettled, however, is the where's, when's, and how's of such investments. While the literature reveals that parental investments are one of the critical inputs in the production of child skills during the first stages of development (see, e.g., Attanasio et al. (2020), List et al. (2018)), evidence also shows that such investments substantially differ across socioeconomic status (SES hereafter; Kalil (2015), Huttenlocher et al. (2010), Hoff (2003b)). Even though the observed SES differences have been consistently observed across space and over time, serving to exacerbate the rising educational and income inequalities that are commonly observed in modern economies, we know little about their underpinnings.

This paper takes a step back from the burgeoning early childhood literature to examine sources of the disparate parental investments and child outcomes across SES in an effort to reveal the potential mechanisms for closing those gaps. We begin by presenting an economic model that invokes parents' beliefs about how parental investments affect child skill formation as a key driver of investments (Cunha et al. (2013)). We then add empirical content to the model by focusing on the first few years of life, when parental investments have been found to play a crucial role (see Fernald et al. (2012), Mistry et al. (2008), Tomopoulos et al. (2006)). To do so, we design two field experiments to explore if changing parental beliefs can be a pathway to improving parental investments in young children. In the first field experiment, over a six-month period starting three days after birth, we use informational nudges informing low-SES parents about skill formation and best practices to foster child development. In the second field experiment, we use a more intensive home visiting program consisting of two visits per month for six months, starting when the child is 24-30 months old.

We operationalize our first field experiment by leveraging the health care system. More specifically, we built partnerships with ten pediatric clinics predominantly serving low-SES families in the Chicagoland area and leveraged the early well-child visits. The intervention is easily replicable and relatively low-cost. In the second field experiment, we provide a home visitation intervention to low-SES families recruited in medical clinics, grocery stores, daycare facilities, community resource fairs, and public transportation in the Chicagoland area. In both cases, we measure the evolution of parents' beliefs, investments and child outcomes at several time points before and after the interventions.

Our analyses point to several unique insights. First, we show that there is a clear SES-gradient in parents' beliefs about the impact of parental investments on child development. A second result provides evidence that these disparities matter, as parents' beliefs predict later cognitive, language, and social-emotional outcomes of their child. For instance, we find that beliefs alone explain up to 18 percent of the observed variation in child language skills. A third insight is that parental beliefs are malleable. Both field experiments induce parents to revise their beliefs. Furthermore, exploiting the random information shocks generated by the experiments, we show that belief revision leads parents to increase investments in their child. For instance, we find that the quality of parent-child interaction is improved after the more intensive intervention (and to a smaller extent, after the less intensive intervention), and we provide evidence of a causal relationship with changes in beliefs about child development. Importantly, the observed impacts on parental investments do not considerably fade for those who participate in the home visiting program (but do fade for those in the lower-intensity field experiment). Finally, we find positive impacts on children's interactions with their parents in both experiments, as well as important improvements in children's vocabulary, math, and social-emotional skills with the home-visiting program months after the end of the intervention. These insights represent a key part of our contribution as they show that changing parental beliefs is a potentially important pathway to improving parental investments and, ultimately, child outcomes.

Beyond gaining direct insights into a key feature underlying observed educational inequities, our work speaks to several branches of the literature. First, it contributes to the literature on parental beliefs by exploring the mutability of such beliefs. Research in Developmental Psychology has found that parental beliefs about child development can predict parenting practices, home environment and child outcomes (Benasich and Brooks-Gunn (1996); Miller (1988)), and also explain socioeconomic disparities in parental language inputs (Rowe (2008, 2018)). Recently, the economics literature introduced parental beliefs about skill formation as a key factor in models of human capital investments (Cunha et al. (2013)). This literature shows that such beliefs differ across SES, and that they predict investments in children (Attanasio et al. (2019); Bhalotra et al. (2020); Biroli et al. (2020); Boneva and Rauh (2018); Carneiro et al. (2019); Cunha et al. (2020)). Not only do we replicate those different findings, but we additionally use two field experiments to demonstrate that parents' beliefs about the impact of parental inputs are malleable and link changes in those beliefs to changes in investments.

A further distinctive aspect of our study is the exploration of the full chain of impacts, from parental beliefs about child development to child outcomes, of two different types of parent-directed interventions. This approach allows us to focus on two types of interventions and to explore if each can change beliefs, and how those belief changes map into parental behaviors. In this way, our data suggest that smaller changes in parental beliefs about child development are not necessarily enough to induce lasting changes in parental investments and child outcomes.

By focusing on parental inputs, we also contribute to the literature on early language interventions in Developmental Psychology (Ferjan Ramírez et al. (2019, 2020); Leech et al. (2018, 2021); Leung et al. (2020); McGillion et al. (2017); Suskind et al. (2016)). These papers show that providing parents with feedback, or coaching, regarding their linguistic inputs can enhance parent-child interactions. Results from our second field experiment confirm those findings in another population, English Language Learners in Spanish-speaking families, and we go one step further by assessing the impact of the intervention on a large spectrum of child outcomes.

The remainder of our paper is organized as follows. Section 2 documents certain facts around the socioeconomic inequalities in parental investments and child outcomes. Section 3 first describes our economic framework that provides a critical link between parental beliefs, parental investments, and child outcomes, and then presents our two field experiments and main findings. Section 4 concludes. In the appendix, we provide additional details on the methodology and results.

2 Disparities in parental investment and child outcomes

The literature has broadly documented important outcome disparities between low and high-SES children (see, e.g., List et al. (2018), Waldfogel and Washbrook (2011)). Of course, given SES is multi-dimensional, there are several ways to show the disparities, but they are broadly consistent with the left-panel of Figure 1, which summarizes one set of results on child outcomes across maternal education bins (from Brooks-Gunn et al. (2006)). The panel shows the dramatic differences in standardized test scores across those born to mothers with various educational levels. The differences are apparent as early as age 3, where those children with mothers who are college graduates are more than one standard deviation above those children who have a mother with a high school degree or lower. The differences maintain and even slightly increase for 18-year-olds.

The right panel in Figure 1 summarizes the differences of one specific parental input (parent/child verbal interaction), though again, the results are broadly consonant with other metrics (see Kalil (2015) and Hoff (2003a) for reviews). In this example, taken from Hart and Risley (1995), the number of words spoken to children varies dramatically across education (income) groups. As the right panel in Figure 1 shows, the verbal communication difference arises early, and expands dramatically by the child's fourth birthday—this is the famous "Thirty Million Word Gap" study.

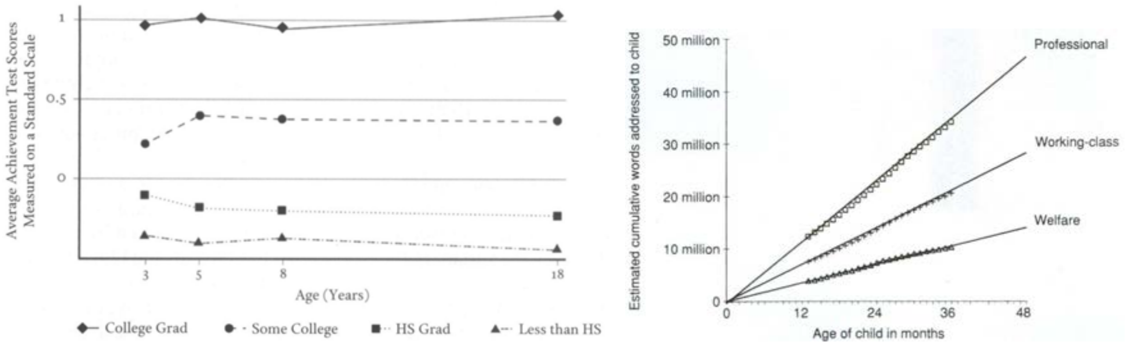


Figure 1: Socioeconomic disparities in parental investments and child outcomes

3 Exploring the roots of investment and outcome disparities

As the literature and the above section highlights, SES disparities permeate child outcomes and parental investments. Understanding their origins and what mechanisms help to attenuate such disparities is key to designing effective policies. While scientists have made important strides to explore the prevalence and importance of those gaps, as shown above, understanding how they arise and what economic environments attenuate or exacerbate their prevalence is an underexplored area of research. We begin with an economic model summarized in section 3.1. The role of the model is to highlight the importance of parental beliefs in the investment process.

To add empirical content to the model, we then explore whether beliefs follow a similar trajectory as parental investments and child outcomes, as observed in Figure 1. Finally, the experimental component must address whether parental beliefs are liable to change and, if so, how changes in beliefs map into parental investment and child outcomes.

3.1 Economic framework

The goal of our set of experiments is to assess the role played by parental beliefs in the development of children's skills. In this section, we present the hypotheses we test within the production technology framework developed by Cunha and Heckman (2007). In their model, the law of motion for skill formation follows some function that depends on the stock of skills and parental investments at the previous period, θ_t and I_t (with t referring to children's age - in our case $t \in \{6, 9, 12, 18\}$ months for the newborn experiment and $t \in \{30-36, 36-42\}$ for the home-visiting experiment), and on parental characteristics (e.g., education, IQ, race, etc.), that we denote X:[1]

$$\theta_{t+1} = f_t(\theta_t, I_t, X)$$

[1] We do not distinguish between cognitive and noncognitive skills here and keep other types of inputs implicit, for simplicity.

In our framework, instead of taking parental investments as given, we go one step further in the understanding of the primitives of early human capital formation and model parents' investments as a function of their contemporaneous and past beliefs about child development, ϕ_t and $\phi_{t-\ell}$, as well as their characteristics and child's skills:

$$I_t = g_t(\phi_t, \phi_{t-\ell}, \theta_t, X)$$

The experiments we designed generate an information shock and are therefore expected to generate exogenous variation in parental beliefs, allowing us to test the sign of the partial derivatives:

$$\frac{\partial I_t}{\partial \phi_t} \gtreqless 0 \qquad \text{and} \qquad \frac{\partial I_t}{\partial \phi_{t-\ell}} \gtreqless 0$$

The closest paper including parents' beliefs in the production function of children's skills is Boneva and Rauh (2017). Combining survey data eliciting beliefs about the returns to parental investments with the British Cohort Study data, the authors estimate a structural model of skill production and simulate the effects of a shift in parental beliefs on investments and long-run economic outcomes of the child. In this paper, we take a reduced-form approach that exploits the randomly assigned treatment as an instrument for beliefs in a Two-Stage Least Squares strategy (2SLS). After checking that the treatment indeed induces a shift in beliefs in the first stage (checking that the instrument is not weak), we regress parents' investments on the predicted beliefs. Results are discussed in section 3.3.

3.2 Two field experiments

The inspiration for our field experiments is contained in the economic model: by changing parents' beliefs about their investment's impact on child development, we can raise their investments, and ultimately improve the child's outcomes. In this manner, to explore the mutability of parental beliefs and their mapping into investments and child outcomes, we designed two framed field experiments targeting low-SES parents of young children ("framed field experiments" is in the parlance of Harrison and List (2004)). As aforementioned, both field experiments took place in the Chicagoland area in the years 2016-2019.

Newborn field experiment

In the first field experiment, which we refer to as the "Newborn program", we partnered with ten pediatric clinics serving medically underserved, underinsured, or uninsured populations in the Chicagoland area (see clinicaltrials.gov, reference NCT02812017). Upon their arrival at the clinic, we worked with every parent during their first well-child visit (3-5 days after birth) on the days and times our research team was present at the clinics. In total, 475 parent-child dyads meeting our socioeconomic and health eligibility criteria were recruited and randomized into the treatment group (237 parents) or control group (238 parents). We obtained consent from the parents only (not the children). More detailed explanations about the recruitment and characteristics of the sample can be found in Appendix B.

Treatment group parents were asked to watch a series of four videos of approximately 10 minutes when they arrived at the clinic for the well-child visits at 1, 2, 4, and 6 months, which correspond to the immunization visits. In the control group, half of the parents were randomly allocated to a placebo intervention that consisted of watching a series of four videos about safety tips for babies at the same well-child visits (see Appendix B for more details). The other half did not watch any video. The two subgroups are pooled together for the main analysis and additional analyses separating the two subgroups are presented in Appendix F.2. Parents would typically watch the video in the waiting room before their appointment with the pediatrician on a tablet provided by a research assistant.

The treatment videos had two main parts: the first provides information on the role of parents in the very early stages of child development and on brain malleability/babies' capabilities, and the second provides practical tips that parents can apply in their daily routines with their baby. The emphasis of the tips was on nurturing language-rich serve-and-return interactions, which we coined as the "3 T's" formula: Tuning in, Talking more, and Taking turns. Such responsive parenting has been shown to be associated with enhanced child vocabulary and overall child development (Gilkerson et al. (2018); Hoff (2003b); Huttenlocher et al. (1991); Rowe (2008); Weisleder and Fernald (2013)). In practice, two-thirds of the parents assigned to the treatment group watched the four videos, one-fourth watched only three videos, and around 8% watched two videos or less. Our intent-to-treat estimates will include all parents, irrespective of the number of videos they watched.

Home-visiting field experiment

The second field experiment, which we denote as the Home Visiting (HV) program, is a more intensive intervention, but it shares the Newborn program's focus on responsive parenting. Like the Newborn program, the HV program emphasizes promoting nurturing language-rich serve-and-return interactions and utilizes the 3Ts framework. For this more intensive program, we recruited parents of 24 to 30 month-old children in medical clinics, grocery stores, daycare facilities, community resource fairs, and public transportation in the Chicagoland area (see clinicaltrials.gov, reference NCT03076268). Among families who consented to participate in the experiment, 91 parent-child dyads met our socioeconomic and health eligibility criteria and were randomized into treatment (46) or control (45) groups (see Appendix B for a more detailed presentation of our sample). Note that we obtained consent from the parents only (not the children).

Parents assigned to the treatment group received two home visits a month for six months. Each of the 12 visits was approximately an hour long, and followed a specific curriculum designed to foster the cognitive and social-emotional development of the child by improving parents' beliefs and investments. During each visit, the home visitor would first show parents a video that covered a specific development topic (e.g., linguistic interactions, encouragement, incorporation of math into everyday routines) and would then do an activity with the caregiver to demonstrate how to put the concepts covered in the video into practice using the 3Ts as a framework. For example, they might practice "tuning in, taking turns, and talking more" about cooking a meal, demonstrating how that daily routine presents a perfect opportunity to engage with a child and introduce descriptive language and math terms. The second part of the visit consisted of providing feedback and setting goals for the next visit in terms of linguistic interactions between the caregiver and the child. Feedback and goals were based on recordings of the child's language exposure and production on a typical day via the LENA technology (see Appendix C). More than two-thirds of the treatment group parents went through the full series of 12 visits, and 24% had less than 6 visits (including no visit at all for 9% of the treatment group). Here again, we will present the intent-to-treat estimates.

To neutralize possible attentional effects (e.g., experimenter demand, Hawthorne effects, accountability, expression of concern from an authority), control group parents received a nutrition intervention. It consisted of sending them packets with information about the importance of healthy nutrition for child development, strategies for healthy eating, and meal preparation. Those packets were reviewed with the parents during short home visits taking place every six weeks over the course of the six-month intervention phase of the study.

In both field experiments, we collected measures of parents' beliefs about their investment's impact on child development, measures of the quality of parent-child interactions and measures of children's skills at regular time points pre- and post-intervention (see the timeline of the data collection for each experiment in Appendix B.2 and the description of the measurement tools in Appendix C).

3.3 Empirical results

Our empirical analysis begins with an exploration of parental beliefs about child development at birth. Since our two field experiments focus on low-SES families, we augment our data with data gathered contemporaneously from a companion study (Suskind et al. (2018)) across various SES levels. Figure 2 summarizes our first result. The left panel compares the belief distribution of low-SES, mid-SES, and high-SES parents. The probability density functions are estimated using the kernel method. Low-SES is defined as household income below 200% of the Federal Poverty Line and the participating parent having at most a bachelors' degree. There are 235 parents in that category in our study. High-SES parents are defined as parents whose household income is above 400% of the Federal Poverty Line (128 observations), and mid-SES parents are neither low- nor high-SES parents (114 observations). The right panel of Figure 2 shows the belief distributions across the educational attainment of the mother. Fifty-two mothers have less than a high-school diploma, 82 have a high-school diploma as their highest degree, 114 have some college education, and 231 have college degrees and beyond.

A first stark result that arises concerns the relationship between parents' socioeconomic status and their beliefs. Both figures paint a consistent and compelling picture: parents with higher levels of education or those from higher socioeconomic backgrounds were more knowledgeable about how parental investments affect child development. For example, in the left panel, the blue and green distributions, respectively, are shifted to the right of the red distribution, suggesting that high-SES and middle-SES parents are more likely to believe that parental investments affect child development than parents of lower SES. Indeed, low-SES parents' beliefs are on average half a standard deviation lower than those of higher SES parents, and they tend to have higher variance (using a Wilcoxon test, we find that low-SES parents have a statistically different distribution of beliefs than both middle and high-SES parents at the 1% level (p-value<0.001 in both cases)). Similarly, the right-hand panel of Figure 2 shows that those mothers with a college education believed parental investments affect child development more than mothers without a college education, and the differences are again significant at the 1% level with p-values<0.001 (comparing parents who have a college degree and beyond with parents with some college education, parents with a high-school degree and parents without a high-school degree).

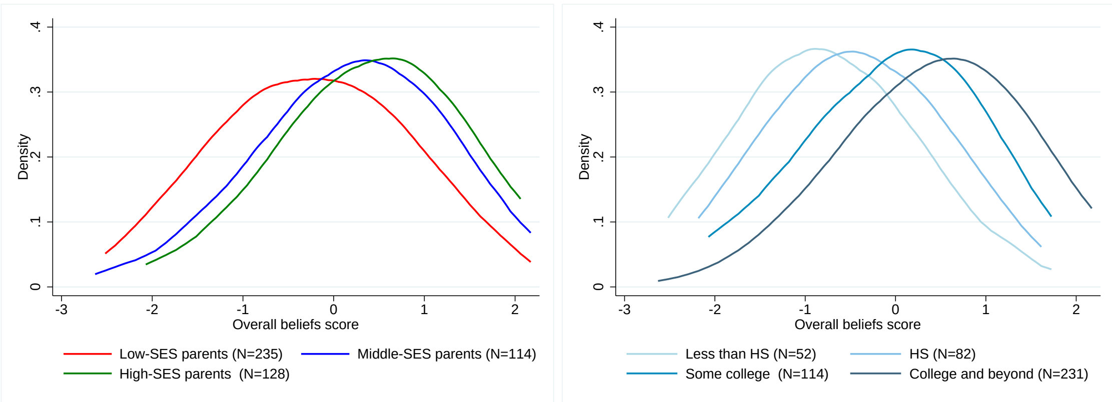


Figure 2: Socioeconomic disparities in parental beliefs about how parental investments affect child skill formation

The socioeconomic gaps in parental beliefs in Figure 2 provide insights into a potential driver of the disparities highlighted in Figure 1. They also provide a key insight that socioeconomic gaps in parental beliefs about child development open up very early in life, much like disparities in children's outcomes (see Bradbury et al. (2015), Reardon (2011), Heckman (2011), Halle et al. (2009)). Investigating the mapping between parents' beliefs, investments, and child development is a critical step to understand how to better address the issue of inter-generational transmission of poverty, which we turn to next.

Field experimental treatment effects

Table 1 summarizes the impacts of our interventions measured on various dimensions at different time points. We display the effect sizes of the comparison of the treatment and control groups, the associated per comparison p-value adjusted for multiple hypotheses testing (see Appendix D for a detailed presentation of our econometric model), and the number of observations used in the regression.

Experimental results show that, even though parental beliefs at baseline were comparable across treatment and control (see Table A.1, Appendix B.1), both field experimental interventions have an immediate and lasting positive impact on parents' beliefs.[2] For example, the top line in Table 1 shows that within 6 months, the treated group in the Newborn program has considerably different beliefs about their investment's impact on child development compared to the control group, a difference that is significant at the 1% level. This statistically significant difference maintains in every time period for both the Newborn program and the HV program. Interestingly, the magnitude of the impact is approximately twice larger in the HV program compared to the Newborn program, and in both cases, tends to decline over time (for Newborn from 0.8σ at the first assessment to 0.6σ in the first six months after the end of the intervention, and 0.4σ after a year). Yet, its statistical significance remains at every time point.
A concern one may have is that these findings partly reflect an experimenter demand effect. We believe that the persistence of the impacts over time, the absence of changes in beliefs in the control group in both experiments, and the assessment protocol mitigate such concern (see Appendix H for a more extensive discussion).

To explore parental investments, we use detailed measures of the quality of parent-child interactions based on audio and video recordings (see Appendix C). As we are studying very young children (infants and toddlers), those measures allow us to capture subtle changes in parents' and children's behaviors that other coarser investment measures commonly used in surveys (e.g., number of books or toys at home, time spent reading or playing with the child) may miss. Moreover, our measures of parental inputs are based on direct observations and blind coding, which protect them against potentially non-random measurement errors that could arise in self-reported survey measures (due to experimenter demand or halo effects, typically).

Looking at the second row of Table 1, we see that there is some evidence of investment improvements. For example, for the Newborn program, the coefficient on parent/child interactions is always positive and reaches significance at the 5% level at the 9 months assessment. For the HV program, we find further evidence of improvements in the parent/child interactions, as a one-sided alternative is significant at the 5% level at our first measurement time point. Taken together, we interpret these estimates as providing some evidence that our treatments, which are found to affect

[2]Note that we also collected other types of parental beliefs (see Appendix C), and when making multiple testing corrections, we account for those measures as well. Those additional results are presented in Appendix E.

beliefs, also affect parental investment. In Appendix G, we examine further the causal relationship between the changes in parents' beliefs and the changes in their investments using a 2SLS strategy. Our results provide evidence that the improvements we observe in parents' investments are driven by the gain in knowledge from the intervention.

Finally, we explore the impacts of our interventions on child outcomes. Summary results are contained in the bottom panel of Table 1. Empirical results suggest that there is some movement in child outcomes with the Newborn program, but the evidence is much stronger for the HV program. For the Newborn program, the behavioral impacts are null right after the intervention, when the child is 6 months old, but become statistically significant three months later for interactions with parents, when the child is 9 months old (the p-values are around 5% at the family level). Improvements are around $+0.2\sigma$, but they fade by the 12 and 18 months assessments. Children's vocabulary, as measured by the MacArthur-Bates Communicative Development Inventories (see Appendix C), is not affected by the intervention overall. By contrast, with the HV program, children's interactions, vocabulary, math skills, and social-emotional skills are all positively impacted by the treatment. Indeed, for each, we find that even after adjusting for multiple testing, there is a considerable amount of evidence suggesting child outcomes are influenced by our home visiting program. There are several potential explanations for why the HV program moves child outcomes more than the Newborn program, but our preferred interpretation is that HV is a much more intensive program compared to the Newborn program.

In Appendix F.1, we show that the results presented in Table 1 are robust to the inclusion of control variables in the estimation.

Table 1: Impacts of the two experiments at different time points

	Newborn program				HV program	
	6 m	7/9 m	12 m	18 m	30-36 m	36-42 m
Parents' beliefs on:						
Impact of parents' inputs	0.82***	0.61***	0.61***	0.38***	1.46***	1.25***
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
	385	284	375	323	69	68
Parents' inputs:						
Interactions w/ child°	0.13	0.27**	0.10	0.10	0.62*	0.44
	0.42	*0.03*	*0.37*	*0.41*	*0.08*	*0.25*
	363	341	374	322	61	60
Child outcomes:						
Interactions w/ parent°	-0.05	0.23**	-0.13	0.13	0.50*	0.30
	0.62	*0.05*	*0.26*	*0.33*	*0.08*	*0.33*
	363	341	374	322	61	60
Vocabulary°		0.09	0.11	-0.04		0.31*
		0.35	*0.45*	*0.58*		*0.08*
		319	369	319		61
Math skills						0.48*
						0.05
						66
Social-emotional skills					0.44**	0.45*

Continued on next page...

... table 1 continued

	Newborn program					
	6 m	7/9 m	12 m	18 m	30-36 m	36-42 m
					0.02	*0.06*
					69	68

Note: Differences between treatment and control group means are shown first for the two experiments at different time points. Below in italics are the multiplicity-adjusted p-values based on a two-sided Student's t test. The adjustment procedure follows List, Shaikh and Xu (2019). Families of outcomes for the adjustment are described in Appendix D. Below the p-values is the number of observations used in the regression. ° signals that the measurement tools used in the two experiments are different (see Appendix C). In the 7/9 months column, beliefs are measured at 7 months, and interactions are measured at 9 months. * for multiplicity-adjusted p-value<0.1, ** for multiplicity-adjusted p-value<0.05, *** for multiplicity-adjusted p-value<0.01.

As previously mentioned, for the Newborn program, Table 1 shows the results of the comparison between the intervention group and the two control subgroups combined (safety video and no video at all). Separating the two control subgroups confirms that the impacts on parental beliefs and inputs presented in Table 1 are driven by the content of the intervention video. The only outcome significantly affected by the safety video is the measure of children's interactions with their parents at 9 months. Both the intervention video and the placebo safety video positively improve children's interactions, leading the coefficient presented in Table 1 to be lower than when the intervention video is compared to no video at all (see the detailed results of the separate comparisons in Appendix F.2).

Beliefs' predictive power

We close our analysis with an exploration of the predictive power of parental beliefs. To do so, we augment our data with those from a companion study that follow parents and their child over a period of two years, starting when the child is 13 to 16 months old, and allow us to study the relationship between parental beliefs and child outcomes longitudinally. The study also started in 2016 among low-SES parents from the Chicagoland area (see Leung et al. (2020) and clinicaltrials.gov, reference NCT02216032). The data contain the same measure of social-emotional skills as in the HV program, as well as two measures of language skills – the Preschool Language Scale and the Peabody Picture Vocabulary Test (see Appendix C). Those skills were regularly assessed over the two years, along with parents' beliefs about the impact of their investments on their child's development (SPEAK survey, see Appendix C).

Table 2 presents the results from simple linear regressions of child skill measures on parental beliefs. For each regression, we present the coefficient of the belief measure, its standard error in parentheses below, the R-squared in italics, and the number of observations. One remarkable result is that the correlations between parental beliefs and children's skills are all positive and mostly significant at 1%, both across different ages and across different skill measures. The strongest correlations are for linguistic skills (Preschool Language Scale (PLS) and Peabody Picture Vocabulary Test (PPVT)), which is not surprising given the emphasis on the role of parent talk in our belief survey.

Empirical results shown in the first column of Table 2 indicate that a 1σ increase in parental knowledge when the child is between 13 and 16 months old is associated with a 0.8 to 1.2σ increase in language skills later in life. The next columns reveal that this relationship can be even stronger at later stages, when the child is 19-22 to 37-40 months old. Considering the R-squared value, we find it important that beliefs alone explain up to 18.7% of the variation in child language skills (for beliefs measured at 19 to 22 months and language skills measured a year and a half later with the PLS) and around 13-15% for most other ages as well as for our other measure of language skills that focuses on vocabulary (PPVT, last row of the table).

Table 2: Predictive power of parental beliefs about child development for children's skills

	Beliefs on impact of parents' inputs at:							
	13-16m		19-22m		25-28m		37-40m	
ASQ-SE at:								
19-22m	0.19*	*4.74%*	0.17*	*3.87%*				
	(0.10)	68	(0.10)	77				
37-40m	0.14	*0.99%*	0.29*	*4.23%*	0.31*	*3.86%*	0.40**	*7.76%*
	(0.18)	67	(0.16)	74	(0.18)	73	(0.16)	76
PLS at:								
13-16m	0.20**	*5.14%*						
	(0.10)	83						
19-22m	0.84***	*13.15%*	0.99***	*15.35%*				
	(0.27)	68	(0.27)	77				
25-28m	1.05***	*13.24%*	1.24***	*15.69%*	1.09***	*9.82%*		
	(0.33)	68	(0.34)	76	(0.39)	75		
31-34m	0.94***	*12.22%*	1.41***	*18.19%*	1.35***	*13.60%*		
	(0.31)	69	(0.35)	76	(0.40)	75		
37-40m	1.24***	*14.40%*	1.40***	*18.65%*	1.40***	*15.03%*	1.25***	*14.09%*
	(0.38)	64	(0.35)	72	(0.40)	70	(0.37)	73
PPVT at:								
37-40m	0.32***	*14.02%*	0.35***	*15.98%*	0.35***	*12.67%*	0.31***	*11.66%*
	(0.10)	63	(0.10)	71	(0.11)	69	(0.10)	72

Note: The rows correspond to children's skills measured at five different time points. ASQ-SE stands for Ages and Stages Questionnaire: Social Emotional, PLS stands for Preschool Language Scale, and PPVT stands for Peabody Picture Vocabulary Test (see Appendix C). They are regressed on parents' beliefs measured at four different time points, in separate simple linear regressions. Both the child outcome variables and the beliefs variables are standardized. For each regression, the table gives the coefficient of the beliefs measure, its standard error in parentheses below, the R-squared in italics next to the coefficient, and the number of observations below the R-squared. * for p-value<0.1, ** for p-value<0.05, *** for p-value<0.01. P-values are based on two-sided Student's t tests without adjustment for multiple comparisons.

4 Conclusions

The literature on child development has taught us that early childhood investments map into long term child outcomes (Campbell et al. (2014); Gertler et al. (2014)), calling for more research on the optimal policies required to reduce early inequities within modern societies. Yet, there is much heterogeneity in childhood investment and child outcomes observed across socioeconomic strata. We approach early childhood disparities differently in that we take a step in the opposite direction from most and ask a basic question: what are the key drivers of parental investment? And, can these drivers help to uncover why we observe constant differences across socioeconomic strata?

Through a simple economic framework, we are directed to focus on parental beliefs about child development. Such an approach proves rich in that it ties parental beliefs to parental investments and, ultimately, to child outcomes. To add empirical content to the model, we change beliefs via two field experiments, and show how these changes in beliefs can be mapped into changes in parental investments. Importantly, we find improvements in various school readiness outcomes with our more intensive program. Though our results point to one important driver of disparate outcomes, we recognize that changing parental beliefs will not address or overcome many of the other deep, structural drivers of inequality.

Nevertheless, our results have several potential policy implications since our approach has various features that are scalable (Al-Ubaydli et al. (2020); List (2020)). They first suggest that providing information and guidance that can change parental beliefs about the impact of parental investments in children can be one pathway to improving school readiness outcomes among low-SES families. They also suggest that simple educational policies may not be sufficient to induce robust behavioral changes and child outcome improvements. Indeed, while both of our interventions show that parental beliefs are malleable, the more intensive program has roughly twice the impact on beliefs as our less intensive intervention. It is a question for future research to understand whether the associated differences in parental investments and child outcomes are due to the level of belief change each program caused. We view our approach as a starting point, which future programs meant to reduce informational frictions can build on to reduce socioeconomic gaps in child development.

Data availability

The datasets generated and/or analyzed during the current study are not publicly available due to the presence of Protected Health Information, but all de-identified datasets will be made available upon request from the corresponding author to replicate the results.

Code availability

All codes used to generate the results presented in this paper are available from the corresponding author on request. All data were collected using REDCap and analyzed using Stata 15.

References

Al-Ubaydli, O., J. List, and D. Suskind (2020). The science of using science: Toward an understanding of the threats to scalability. *International Economic Review 61*(4), 1387–1409.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2020). Estimating the production function for human capital: results from a randomized controlled trial in Colombia. *American Economic Review 110*(1), 48–85.

Attanasio, O., F. Cunha, and P. Jervis (2019). Subjective parental beliefs. their measurement and role. Technical report, National Bureau of Economic Research.

Becker, G. S. (1993). *Human capital: A theoretical and empirical analysis, with special reference to education.* University of Chicago Press.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies 81*(2), 608–650.

Benasich, A. A. and J. Brooks-Gunn (1996). Maternal attitudes and knowledge of child-rearing: Associations with family and child outcomes. *Child Development 67*(3), 1186–1205.

Bhalotra, S. R., A. Delavande, P. Gilabert, and J. Maselko (2020). Maternal investments in children: The role of expected effort and returns. *IZA Discussion Paper*.

Biroli, P., T. Boneva, A. Raja, and C. Rauh (2020). Parental beliefs about returns to child health investments. *Journal of Econometrics (in press).*

Bloomfield, L. and S. Kendall (2010). Audit as evidence: the effectiveness of '123 magic' programmes. *Community Practitioner*.

Boltho, A. and G. Toniolo (1999). The assessment: the twentieth century. *Oxford Review of Economic Policy 15*(4), 1–17.

Boneva, T. and C. Rauh (2017). Human capital production and parental beliefs. *Unpublished manuscript*.

Boneva, T. and C. Rauh (2018). Parental beliefs about returns to educational investments—the later the better? *Journal of the European Economic Association 16*(6), 1669–1711.

Bradbury, B., M. Corak, J. Waldfogel, and E. Washbrook (2015). *Too many children left behind: The US achievement gap in comparative perspective.* Russell Sage Foundation.

Brooks-Gunn, J., F. Cunha, G. Duncan, J. J. Heckman, and A. Sojourner (2006). A reanalysis of the ihdp program. *Unpublished manuscript, Infant Health and Development Program, Northwestern University.*

Brownell, R. (2010). *Receptive and expressive one-word picture vocabulary tests.* Bloomington, MN: Pearson Assessments.

Campbell, F., G. Conti, J. J. Heckman, S. H. Moon, R. Pinto, E. Pungello, and Y. Pan (2014). Early childhood investments substantially boost adult health. *Science 343*(6178), 1478–1485.

Carneiro, P. M., E. Galasso, I. X. Lopez Garcia, P. Bedregal, and M. Cordero (2019). Parental beliefs, investments, and child development: Evidence from a large-scale experiment. *World Bank Policy Research Working Paper* (8743).

Cunha, F., I. Elo, and J. Culhane (2013). Eliciting maternal expectations about the technology of cognitive skill formation. Technical report, National Bureau of Economic Research.

Cunha, F., I. Elo, and J. Culhane (2020). Maternal subjective expectations about the technology of skill formation predict investments in children one year later. *Journal of Econometrics (in press).*

Cunha, F. and J. Heckman (2007). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Dunn, L. M. and D. M. Dunn (2007). *Peabody Picture Vocabulary Test, Fourth Edition.* Pearson Education.

Dweck, C. S. (2013). *Self-theories: Their role in motivation, personality, and development.* Psychology Press.

Fenson, L., S. Pethick, C. Renda, J. L. Cox, P. S. Dale, and J. S. Reznick (2000). Short-form versions of the macarthur communicative development inventories. *Applied Psycholinguistics 21*(1), 95–116.

Ferjan Ramírez, N., S. R. Lytle, M. Fish, and P. K. Kuhl (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., S. R. Lytle, and P. K. Kuhl (2020). Parent coaching increases conversational turns and advances infant language development. *Proceedings of the National Academy of Sciences 117*(7), 3484–3491.

Fernald, L. C., P. Kariger, M. Hidrobo, and P. J. Gertler (2012). Socioeconomic gradients in child development in very young children: Evidence from India, Indonesia, Peru, and Senegal. *Proceedings of the National Academy of Sciences 109*(Supplement 2), 17273–17280.

Gertler, P., J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S. M. Chang, and S. Grantham-McGregor (2014). Labor market returns to an early childhood stimulation intervention in Jamaica. *Science 344*(6187), 998–1001.

Gilkerson, J., J. A. Richards, S. F. Warren, D. K. Oller, R. Russo, and B. Vohr (2018). Language experience in the second year of life and language outcomes in late childhood. *Pediatrics 142*(4), e20174276.

Halle, T., N. Forry, E. Hair, K. Perper, L. Wandner, J. Wessel, and J. Vick (2009). Disparities in early learning and development: lessons from the early childhood longitudinal study–birth cohort (ecls-b). *Washington, DC: Child Trends*, 1–7.

Harrison, G. W. and J. A. List (2004). Field experiments. *Journal of Economic Literature 42*(4), 1009–1055.

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

Heckman, J. J. (2011). The American family in Black & White: A post-racial strategy for improving skills to promote equality. *Daedalus 140*(2), 70–89.

Heckman, J. J. (2012). Invest in early childhood development: Reduce deficits, strengthen the economy. *The Heckman Equation 7*, 1–2.

Hoff, E. (2003a). Causes and consequences of ses-related differences in parent-to-child speech. In *Monographs in parenting series. Socioeconomic status, parenting, and child development*, pp. 147–160. Lawrence Erlbaum Associates Publishers.

Hoff, E. (2003b). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development 74*(5), 1368–1378.

Huttenlocher, J., W. Haight, A. Bryk, M. Seltzer, and T. Lyons (1991). Early vocabulary growth: relation to language input and gender. *Developmental Psychology 27*(2), 236.

Huttenlocher, J., H. Waterfall, M. Vasilyeva, J. Vevea, and L. V. Hedges (2010). Sources of variability in children's language growth. *Cognitive psychology 61*(4), 343–365.

Kalil, A. (2015). Inequality begins at home: The role of parenting in the diverging destinies of rich and poor children. In *Families in an era of increasing inequality*, pp. 63–82. Springer.

Kendall, S. and L. Bloomfield (2005). Developing and validating a tool to measure parenting self-efficacy. *Journal of Advanced Nursing 51*(2), 174–181.

Kolobe, T. H. (2004). Childrearing practices and developmental expectations for mexican-american mothers and the developmental status of their infants. *Physical Therapy 84*(5), 439–453.

Leech, K., R. Wei, J. R. Harring, and M. L. Rowe (2018). A brief parent-focused intervention to improve preschoolers' conversational skills and school readiness. *Developmental Psychology 54*(1), 15.

Leech, K. A. et al. (2021). An intervention to increase conversational turns between parents and young children. *Journal of Child Language 48*(2), 399–412.

Leung, C. Y., M. W. Hernandez, and D. L. Suskind (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.

List, J. A. (2020). Non est disputandum de generalizability? A glimpse into the external validity trial. Technical report, National Bureau of Economic Research.

List, J. A., J. Pernaudet, and D. L. Suskind (2021). Shifting parental beliefs about child development to foster parental investments and improve school readiness outcomes. *Nature Communications 12*(1), 1–10.

List, J. A., A. Samek, and D. L. Suskind (2018). Combining behavioral economics and field experiments to reimagine early childhood education. *Behavioural Public Policy 2*(1), 1–21.

List, J. A., A. M. Shaikh, and Y. Xu (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics 22*(4), 773–793.

Lucas, R. (1988). On the mechanics of economic development. *Journal of Monetary Economics 22*, 3–42.

Lugo-Gil, J. and C. S. Tamis-LeMonda (2008). Family resources and parenting quality: Links to children's cognitive development across the first 3 years. *Child Development 79*(4), 1065–1085.

McGillion, M., J. M. Pine, J. S. Herbert, and D. Matthews (2017). A randomised 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.

Miller, S. A. (1988). Parents' beliefs about children's cognitive development. *Child Development*, 259–285.

Mistry, R. S., J. C. Biesanz, N. Chien, C. Howes, and A. D. Benner (2008). Socioeconomic status, parental investments, and the cognitive and behavioral outcomes of low-income children from immigrant and native households. *Early Childhood Research Quarterly 23*(2), 193–212.

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. *Whither Opportunity 1*(1), 91–116.

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of child language 35*(1), 185–205.

Rowe, M. L. (2018). Understanding socioeconomic differences in parents' speech to children. *Child Development Perspectives 12*(2), 122–127.

Squires, J., D. Bricker, and E. Twombly (2002). *The ASQ: SE user's guide: For the Ages & Stages Questionnaires: Social-emotional.* Paul H Brookes Publishing.

Sumner, G. and A. Spietz (1994). NCAST caregiver/parent-child interaction teaching manual. Seattle. *NCAST Publications*.

Suskind, D. L., K. R. Leffel, E. Graf, M. W. Hernandez, E. A. Gunderson, S. G. Sapolich, E. Suskind, L. Leininger, S. Goldin-Meadow, and S. C. Levine (2016). A parent-directed language intervention for children of low socioeconomic status: A randomized controlled pilot study. *Journal of child language 43*(2), 366–406.

Suskind, D. L., C. Y. Leung, R. J. Webber, A. C. Hundertmark, K. R. Leffel, I. E. Fuenmayor Rivas, and W. A. Grobman (2018). Educating parents about infant language development: A randomized controlled trial. *Clinical Pediatrics 57*(8), 945–953.

Suskind, D. L., C. Y. Leung, R. J. Webber, A. C. Hundertmark, K. R. Leffel, E. Suskind, M. W. Hernandez, and E. Graf (2018). Development of the Survey of Parent/Provider Expectations and Knowledge (SPEAK). *First Language 38*(3), 312–331.

Tomopoulos, S., B. P. Dreyer, C. Tamis-LeMonda, V. Flynn, I. Rovira, W. Tineo, and A. L. Mendelsohn (2006). Books, toys, parent-child interaction, and development in young latino children. *Ambulatory Pediatrics 6*(2), 72–78.

Waldfogel, J. and E. Washbrook (2011). Early years policy. *Child Development Research*.

Weisleder, A. and A. Fernald (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological science 24*(11), 2143–2152.

Woodcock, R. W., A. F. Muñoz-Sandoval, M. L. Ruef, and C. G. Alvarado (2005). *Woodcock-Muñoz Language Survey–Revised*. Itasca, IL: Riverside.

Zimmerman, I. L., V. G. Steiner, and Pond (2013). *PLS-5: Preschool Language Scales*. Pearson.

Appendix

A Reporting checklist

A shorter version of the paper was published by *Nature Communications* (see List et al. (2021)). We followed the checklist that we include below to report the results in the paper. It also includes additional methodological details about the studies and analyses.

nature research

Corresponding author(s): Dana L. Suskind

Last updated by author(s): Jul 7, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	REDCap was used to collect the data for the Newborn and HV studies, and the CRI platform was used for the Longitudinal study (http://cri.uchicago.edu/cri-developing-new-community-platform-with-thirty-million-words/).
Data analysis	The website Research Randomizer was used for the randomization for all studies. Stata 15 was used to analyze the data for all studies. No custom code has been developped for the data analysis.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The datasets generated and/or analyzed during the current study are not publicly available due to the presence of Protected Health Information but all de-identified datasets will be made available upon request from the corresponding author to replicate the results. Source data are provided with this paper.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Quantitative experimental study using longitudinal data on parental beliefs, behaviors, and child outcomes
Research sample	Newborn study: Spanish or English-speaking low-SES parents of newborns recruited in 10 pediatric clinics of the Chicagoland area serving medically underserved, underinsured, or uninsured populations. HV study: Spanish-speaking low-SES parents of 24 to 30 months old children recruited through postings in medical clinics, grocery stores, daycare facilities, community resource fairs, and public transportation in the Chicagoland area. Longitudinal study: English-speaking low-SES parents of 13 to 16 months old children recruited through postings at child care centers, libraries, health clinics, local stores, public transportation, and community organizations serving low-income populations in the Chicagoland area. Detailed recruitment procedures are provided in the Methods section. We focus on low-SES families as there is ample evidence that parental investments and child outcomes are lower among those families, as explained in the paper. A series of demographic characteristics are provided in the Supplementary Information, section 3 to describe the population our samples are representative of in more details.
Sampling strategy	Power calculations were made to determine the sample size. Those calculations are presented in the Methods section. The sampling strategy was designed to enroll low-SES English-speaking and/or Spanish-speaking families living in the Chicagoland area. It was based on voluntary response sampling as explained above.
Data collection	Surveys filled out by parents were used to collect beliefs data (pen and paper or tablets were used at the clinic in the presence of a trained RA or an email was sent to complete it at home). Video and audio recordings using the NCAST and LENA instruments, respectively, were used at the clinic in the presence of a trained RA (NCAST) and at home (LENA) to collect parental inputs. Surveys filled out by parents (pen and paper or tablets at the clinic in the presence of a trained RA) and in-person assessments were used at home in the presence of a trained assessor to collect child outcomes (see Supplementary Table 3 for a complete description). Assessors were not blind to experimental condition and study hypothesis.
Timing	For the newborn study, the first enrollment took place on June 20, 2016 and the last enrollment on July 31, 2017. For the home-visiting study, the first enrollment took place on April 17, 2017 and the last enrollment on July 9, 2018. For each study, only one cohort was enrolled and we collected data at regular time points for that cohort (see the specific timelines in the Methods section), following the same time intervals for all participants (e.g., a participant enrolled in June 2016 in the newborn study would be surveyed again in December 2016 for the 6-month assessment while a participant enrolled in January 2017 would be surveyed again in July 2017).
Data exclusions	No data were excluded from the analyses
Non-participation	Non-participation rates are described in detail in the Methods section.
Randomization	Participants were randomly allocated to groups using Research Randomizer website. Additional details are provided in the Methods section.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
[x]	[ ] Antibodies
[x]	[ ] Eukaryotic cell lines
[x]	[ ] Palaeontology and archaeology
[x]	[ ] Animals and other organisms
[ ]	[x] Human research participants
[ ]	[x] Clinical data
[x]	[ ] Dual use research of concern

Methods

n/a	Involved in the study
[x]	[ ] ChIP-seq
[x]	[ ] Flow cytometry
[x]	[ ] MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	See above
Recruitment	Newborn study: parents were recruited in pediatric clinics by trained RAs. HV study: parents were recruited through postings in medical clinics, grocery stores, daycare facilities, community resource fairs, and public transportation. Longitudinal study: parents were recruited through postings at child care centers, libraries, health clinics, local stores, public transportation, and community organizations. In the three cases, parents had to first volunteer to participate in the study, and then they were screened to check that they were eligible following the criteria described in the Methods section. The first step implies that only parents interested in the topic of the study were ultimately recruited. We do not believe that this self-selection creates any major issue for the interpretation of our results. The second step implies that our sample is representative of a specific population who presents the characteristics we used to determine eligibility. This is not an issue for the interpretation of our results as long as we do not extrapolate to other populations.
Ethics oversight	For all the studies involved in the analysis, we received approval from the University of Chicago Biological Sciences Division Institutional Review Board.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	Newborn study: NCT02812017 HV study: NCT03076268 Longitudinal study: NCT02216032
Study protocol	The full study protocols can be accessed on clinicaltrials.gov: Newborn study: https://clinicaltrials.gov/ct2/show/NCT02812017 HV study: https://clinicaltrials.gov/ct2/show/NCT03076268 Longitudinal study: https://clinicaltrials.gov/ct2/show/NCT02216032
Data collection	See above
Outcomes	Following our theory of change (described in the paper and in the Supplementary Information, section 1), our primary outcomes are parental beliefs, parental investments and child outcomes. We assess those measures comparing treatment and control groups via simple linear regressions.

B Design of the experiments

B.1 Sample, randomization, and balancing checks

In addition to the income and education criteria described in the paper, we applied age, language, and health criteria to select parents. In both experiments, eligible parents had to be female primary caregivers, at least 18 years old, had to live with their child and have legal custody, and their child had to fall into a certain age range (less than 30 days old for the newborn program and 24 to 30 months old for the home-visiting program). In the newborn program, parents had to speak English or Spanish, whereas in the home-visiting program, they had to speak Spanish as their preferred language at home. Finally, children born with significant perinatal or neonatal complications, developmental disabilities or medical problems, and before 36 weeks were excluded from the newborn study, and children with significant cognitive or physical impairments were excluded from the home-visiting study. In both experiments, parents who lived above 200% of the federal poverty line, parents whose education level was more than a bachelor's degree, foster parents, and parents who were unable to commit to the intervention requirements were also excluded. The full protocols are registered on clinicaltrials.gov, references NCT02812017 and NCT03076268.

For all the studies involved in the analysis, we received approval from the University of Chicago Biological Sciences Division Institutional Review Board. The experiments complied with all relevant ethical regulations for work with human participants and involved informed consent of the parents that was obtained through written signature after they read the form and asked any questions they may have to a trained Research Assistant.

For the newborn study, we targeted a baseline sample size of a minimum of 400 participants based on power calculations setting a target power threshold of 80%, a significance level of 5%, and assuming a 25% dropout rate. Data from a previous study by Suskind et al. (2018) were used to assess the size of expected treatment effects on parental beliefs (measured by the Survey of Parents' Expectations And Knowledge, see Appendix C).

For the home-visiting study, we targeted a baseline sample size of a minimum of 90 participants based on power calculations also setting a target power threshold of 80%, a significance level of 5%, and assuming a 25% dropout rate. Data from our Longitudinal home-visiting study (Leung et al. (2020)) were used to assess the size of expected treatment effects on parent-child interactions (measured by the LENA recordings, see Appendix C).

For the newborn study, the first enrollment took place on June 20, 2016 and the last enrollment on July 31, 2017. For the home-visiting study, the first enrollment took place on April 17, 2017 and the last enrollment on July 9, 2018.

To provide a more detailed description of each sample, Table A.1 below shows the average characteristics of the control group measured in our baseline surveys for each study. Participants are by construction from low-SES families: almost a third of them do not have a high-school diploma, around 80% have a household income below $2,655 per month and a large share of families are enrolled in social programs such as insurance or food programs in both samples. In the newborn sample, 35% of the respondents were employed at the time of the recruitment, while this proportion is 51% in the home-visiting sample. 54% and 100%, respectively, are Hispanic or Latino in the newborn and home-visiting study, and a little less than 30% declare themselves White in both cases.

Parents who enrolled in the experiments were randomly assigned to the treatment or control group so that the two groups are on average similar, and a simple comparison of their outcomes post-intervention allows us to recover the causal impact of the intervention. In both experiments, the randomization was done using the website Research Randomizer to generate a list of unique unsorted numbers (91 numbers in the HV program and 475 numbers in the Newborn program), each paired with a participant number. The first half of the random sequence was then assigned to Treatment, and the second half to Control. For the Newborn program, the procedure was done separately for English-speaking and Spanish-speaking participants.

We cannot directly test whether the randomization indeed created similar groups, but we can compare control and treatment parents along their baseline characteristics to check that at least their observed characteristics are on average similar. Those comparisons are shown in the second and fifth columns of Table A.1 (T-C) for the set of demographic variables, a synthetic dummy built from the Family Life Events survey indicating whether the parent reported at least one traumatic event that occurred in the past 6 months (divorce, death, violent crime, serious illness, depression, prison, child lived with someone with alcohol or drug issues, extreme financial difficulties), and the belief scores.

The comparisons reveal that some variables tend to be slightly unbalanced (parent's age and the likelihood of experiencing a traumatic event in the newborn study; employment status in both studies) but the difference is generally significant only at the 10% level without accounting for multiple hypotheses testing, except for the employment status in the home-visiting study which is significantly different at 1%. Since we estimate impacts separately at each time point in the impact analysis and there is attrition between each time point that could generate further imbalance (if related to the treatment status), we did those comparisons separately at each time point (tables available upon request). A few variables become imbalanced at later time points in the newborn study (LINK card, number of children). Regressions taking into account imbalances between the treatment and control groups are presented in Appendix F.1.

Table A.1: Balancing checks at baseline

Variable	Newborn program Mean C	T-C	N	HV program Mean C	T-C	N
1 if child is a girl	0.53	0.00	475	0.44	-0.12	91
		0.96			*0.25*	
Parent's age	26.01	1.01*	475	33.00	0.48	91
		0.06			*0.77*	
Highest educ. level < High-school dip.	0.32	-0.05	475	0.33	0.06	91
		0.24			*0.57*	
Number of children	2.17	0.09	469	2.67	0.05	91
		0.42			*0.85*	
Household income per month < $2 655	0.83	0.01	475	0.80	0.03	91
		0.73			*0.75*	
Primary language is English	0.65	-0.01	469	.	.	.
		0.90			.	
Black or African American	0.42	0.03	469	.	.	.
		0.55			.	
White	0.26	-0.01	469	0.29	0.10	91
		0.77			*0.31*	
Hispanic or Latino	0.54	-0.01	469	1.00	.	91
		0.89			.	
Medicaid/Medical card	0.85	0.03	469	0.96	-0.07	90
		0.35			*0.24*	
LINK card	0.52	0.07	469	0.62	-0.06	91
		0.15			*0.59*	
WIC supplements	0.70	-0.04	469	0.64	-0.12	91
		0.41			*0.24*	
Currently employed	0.35	-0.07*	469	0.51	-0.32***	91
		0.10			*0.00*	
Full-time student	0.05	-0.03	469	0.04	-0.04	91
		0.15			*0.16*	
Most recent job full-time	0.33	0.06	468	0.53	0.02	90
		0.18			*0.84*	
Most recent job part-time	0.37	-0.01	468	0.31	-0.07	90
		0.85			*0.49*	
Recent traumatic event	0.24	0.08*	475	0.44	-0.10	91
		0.06			*0.35*	
Beliefs about impact of parental inputs	43.99	0.69	471	46.50	1.57	82
		0.41			*0.43*	

Note: Mean C gives the mean of the variable in the control group, T-C gives the difference with the treatment group. Below in italics is the corresponding p-value based on a two-sided Student's t test without adjustment for multiple comparisons. N gives the number of observations. * for p-value<0.1, ** for p-value<0.05, *** for p-value<0.01.

B.2 Timeline, enrollment rates and attrition

Figure A.1 below presents the timing of the different surveys parents completed and the videos they were asked to watch for the newborn study (top panel) or the home visits they received in the second study (bottom panel), along with the payment they received to do it. The acronyms of the surveys are explained in section C. In the top panel, the TMW curriculum modules correspond to the videos of the intervention, the safety videos are videos watched by our placebo control group (see section F for explanations about the placebo control group).

Figure A.1: Timeline of the two experiments and surveys

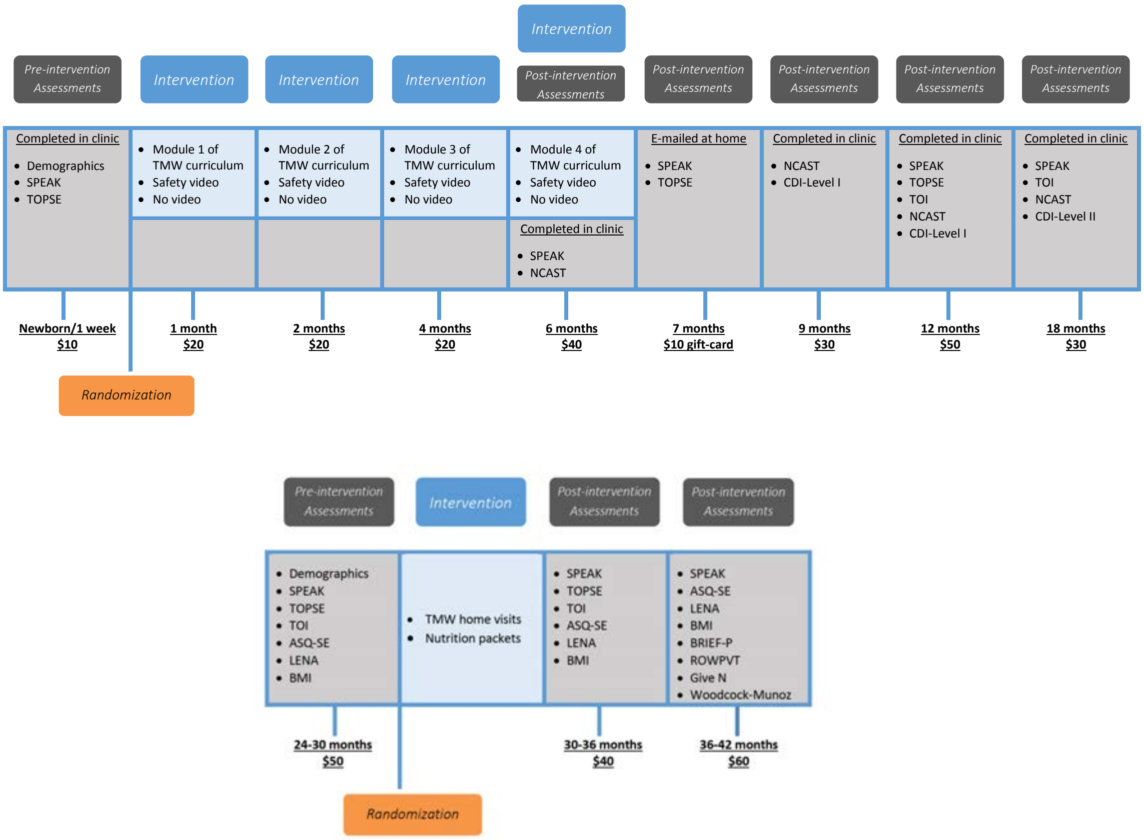
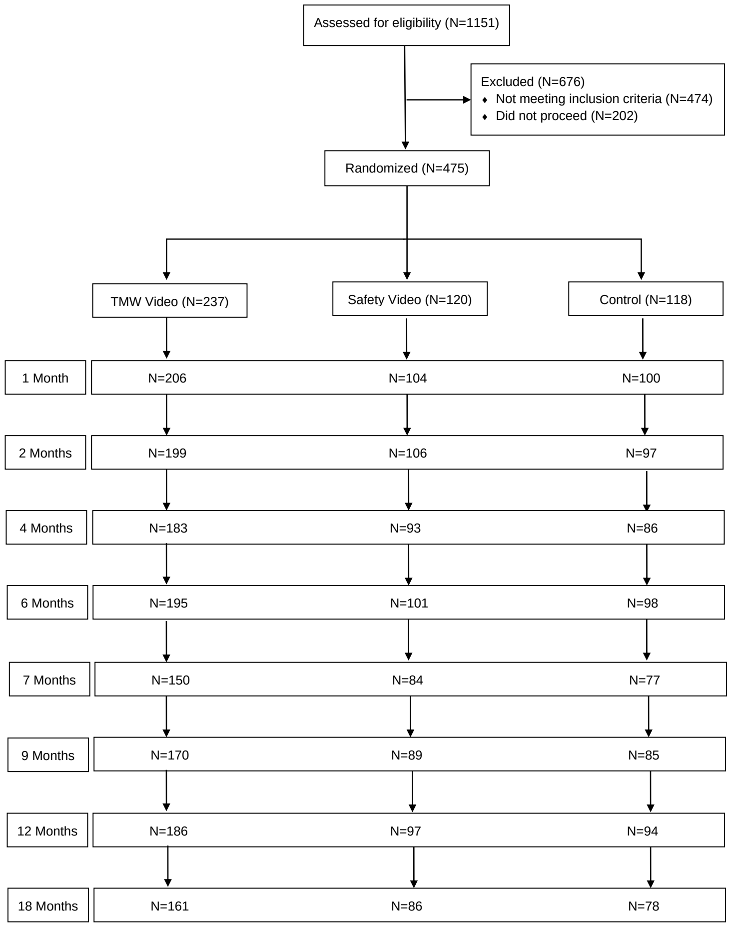
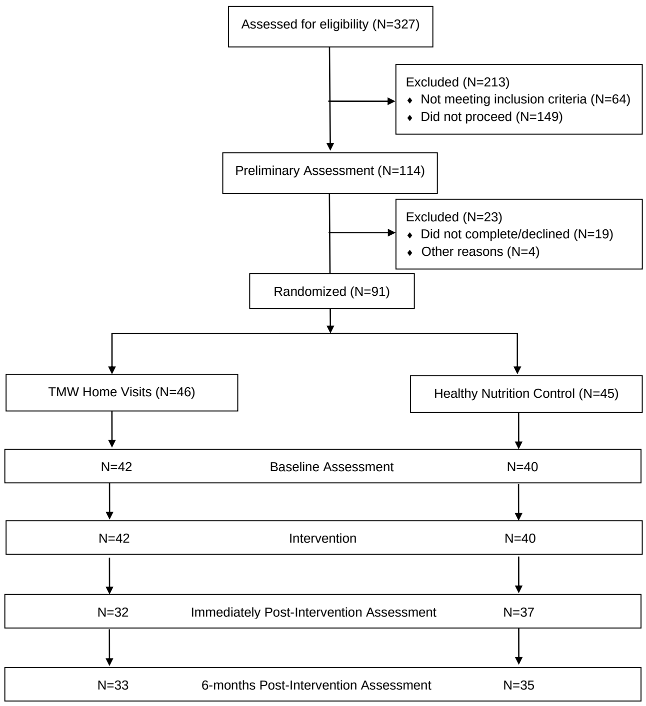


Figure A.2 below gives the number of parents initially approached, the number of parents who were randomized, as well as the attrition over time in each group for both experiments.

Figure A.2: Enrollment and attrition rates at each assessment for the two experiments





C Measurement tools

This section presents the different measurement tools used for the analyses. Table A.2 below first provides a summary of all the variables used in the paper, explaining where they were used in the analysis and how they were measured. We then present in detail the content and format of each measurement tool.

Table A.2: Summary of the measures used in the analysis

Variable	Study	Used in	Measurement
Parents' beliefs:			
SPEAK	- Suskind et al. (2018)	Figure 2	Self-report
	- Newborn	Tables 1, A.7, A.3, A.5, A.6, Fig. A.3	
	- HV	Tables 1, A.7, A.4, A.5, Fig. A.3	
	- Longitudinal	Table 2	
TOPSE	- Newborn	Tables A.3, A.6	Self-report
	- HV	Table A.4	
TOI	- Newborn	Tables A.3, A.6	Self-report
	- HV	Table A.4	
Parents' inputs:			
NCAST	Newborn	Tables 1, A.7, A.3, A.5, A.6	Direct obs.
LENA	HV	Tables 1, A.7, A.5	Direct obs.
Child outcomes:			
CDI I and II	Newborn	Tables 1, A.5, A.6	Self-report
ASQ-SE	- HV	Tables 1, A.5	Self-report
	- Longitudinal	Table 2	
PLS	Longitudinal	Table 2	Direct obs.
Woodcock-Muñoz	HV	Tables 1, A.5	
ROWPVT	HV	Tables 1, A.5	Direct obs.
PPVT	Longitudinal	Table 2	Direct obs.
Give N	HV	Tables 1, A.5	Direct obs.

Note: All variables except Woodcodk-Munoz and ROWPVT were analyzed as overall standardized scores. Standardization was made with respect to the mean and standard deviation of the full baseline sample distribution when the measure was available at baseline, and otherwise with respect to the distribution of the control group only at the earliest assessment point. For Woodcodk-Muñoz and ROWPVT, we built an index combining the two measures (average of the two z-scores).

Note that the full list of primary outcomes that were collected as part of the two experiments can be found on clinicaltrials.gov (references NCT02812017 and NCT03076268). In the Newborn study, we included in the paper all the primary outcomes observed repeatedly across the different assessment points as we are interested in assessing the evolution of the impacts. We thus excluded the StimQ Cognitive Home Environment and Child Behavioral Checklist, measured only once at later assessment points. In the home-visiting study, we included all primary outcomes.

C.1 Parental beliefs

SPEAK survey

The Survey of Parents' Expectations And Knowledge (SPEAK) is the tool we use to elicit parental beliefs about how parents' input impact child development for both the descriptive and impact analyses. Note that throughout the paper, we consider that parental beliefs are knowledge-based (Rowe (2008)) and use the terms 'beliefs' and 'knowledge' interchangeably. The SPEAK survey has been developed by Suskind et al. (2018) and focuses on beliefs about social-emotional and cognitive (with an emphasis on language) development of children between 0 and 5 years old. It consists of a series of questions that cover topics such as the age when babies can start being exposed to different types of reasoning (e.g., counting, sizes, shapes), the effect of media use or bilingualism on learning, the role of nature versus nurture in child early development, and best communication and reading practices. Parents have to answer on a four-point scale from "Definitely True" to "Definitely Not True" (except for questions about stages of development for which they had to select a stage, from 0 to 6 years old and more). The total score is simply the sum of the points over the series of questions. A higher score indicates a better understanding of the impact of parental inputs on child development. Parents complete the survey in English or Spanish, depending on their preferred language. Note that the survey is still in development and the formulation of some of the questions varies slightly between the version used to produce Figure 2 and the version used in the two experiments.
Few measurement tools are available in the literature to assess parental beliefs about child development at early stages. Cunha et al. (2013), Attanasio et al. (2019) and a series of related papers use hypothetical scenarios to elicit parental beliefs about the parametrization of the technology of skill formation. Here, we aim at assessing parental knowledge of early development more generally, covering a broad range of parenting practices that are relevant for the development of children's skills (media use, reading, responsiveness, languages spoken at home, etc.)

TOPSE survey

The second measure of parental beliefs we use is a scale called TOol to measure Parenting Self-Efficacy (TOPSE) that captures parental self-efficacy (Kendall and Bloomfield (2005)). We use the short version of it (Bloomfield and Kendall (2010)). It assesses parents' perception of their own ability to parent effectively in six dimensions: Emotion and Affection (e.g., "I can recognize when my child is happy or sad"), Play and Enjoyment (e.g., "I can plan activities that my child will enjoy"), Empathy and Understanding (e.g., "I am able to put myself in my child's shoes"), Pressure (e.g., "It is difficult to cope with other people's expectations of me as a parent"), Self-Acceptance (e.g., "I can be strong for my child"), and Learning and Knowledge (e.g., "I am able to learn and use new ways of dealing with my child"). Although each subscale can be studied independently, we only look at the overall score in the analysis, to limit the number of hypotheses we test. Note that the tool contains two other subscales, Control, and Discipline and Boundaries, that were not included in our survey as they were less directly relevant to the intervention and the duration of the assessment was constrained. Parents had to answer on a ten-point scale from "Completely Agree" to "Completely Disagree". Each item is scored between 0 and 10 points and the total score is the sum of all items. A higher score indicates higher parenting self-efficacy.

TOI survey

The Theory Of Intelligence (TOI) survey captures parents' beliefs about intelligence malleability of young children in a series of eight items with which parents have to say to what extent they agree, on a five-point Likert scale (Dweck (2013)). It focuses solely on the extent to which children can change their intelligence or ability, not on the elasticity of skills with respect to parental inputs (e.g., "A child's intelligence is something about them that they can't change very much").

C.2 Parent-child interactions

NCAST

To measure parents' inputs in the Newborn study, we rely on a direct assessment of parent-child interaction during a teaching episode, a measurement tool called the Nursing Child Assessment Satellite Training (NCAST) scale (Sumner and Spietz (1994)) that can be used for children as young as six months old. Each parent is filmed in a private room with his or her baby for 6 minutes while teaching a given task (e.g., stacking blocks, drawing). Then the video is coded by a trained external assessor along six subscales describing the quality of the interaction. The first four subscales describe the parent behavior towards the child: his or her sensitivity to cues (e.g., "Caregiver praises child's successes or partial successes"), response to child's distress (e.g., "Caregiver makes a positive, sympathetic, or soothing verbalization"), social-emotional growth fostering (e.g., "Caregivers makes cheerleading type statements to the child during the teaching interaction"), and cognitive growth fostering (e.g., "Caregiver responds to the child's vocalizations with a verbal response"). We sum those four subscales in one single parental score of parent's interactions with the child for the main analysis (disaggregated results are presented in Table A.3). The last two subscales describe the child behavior towards the caregiver: clarity of cues (e.g., "Child makes clearly recognizable arm movement during the teaching episode (clapping, reaching, waving, pointing, ...)") and responsiveness to the caregiver (e.g., "Child vocalizes or babbles within five seconds after caregiver 's verbalization"). Similarly, we sum those two subscales in one single child score of child's interactions with the parent for the main analysis (disaggregated results are presented in Table A.3). For each of the 73 items, the assessor simply chooses between Yes and No, and the score is the sum of Yes answers. Those direct assessments of parent-child interaction give us detailed and objective measures of parental inputs that are relevant for children as young as 6 months old, and that have been found to predict later cognitive and psycho-motor outcomes (Kolobe (2004); Lugo-Gil and Tamis-LeMonda (2008)).

LENA

In the home visiting program, parent-child interactions were measured using the Language ENvironment Analysis (LENA) technology. The measurement consists of using an audio-recording device that the child wears in a specially-designed t-shirt for approximately 8 hours on each recording day. The device records all sounds produced around the child and the data are then processed by a speech recognition software to generate three count-based metrics (see `www.lena.org`). In Table 1, parents' interactions with the child are measured through conversational turn counts (CTC) and child's interactions with the parents are measured through vocalization counts (CVC). We do not use the third metric called the adult word count (AWC) as this metric is not child-directed and counts any adult words around the child, including adults talking on the phone or talking to each other in the room, and does not correspond to an outcome of interest in our study, which focuses on parent-child interactions.

C.3 Children's skills

CDI

In the Newborn experiment, we measure child's vocabulary at 9, 12 and 18 months using the short form of the MacArthur-Bates Communicative Development Inventories (Fenson et al. (2000)) (CDI). The forms are filled out by parents and vary with age. At 9 and 12 months (level I), parents are presented with a list of words and have to indicate, for each word, whether their child understands it (receptive vocabulary) or understands and says it (expressive vocabulary). Both scores have been transformed in z-scores and averaged (disaggregated results are presented in Table A.3). At 18 months (level II), parents are presented with a list of words and have to indicate, for each word, whether their child says it (expressive vocabulary) and they answer a question asking whether their child is capable of combining words. Both questions have been transformed in z-scores and averaged (disaggregated results are presented in Table A.3). Parents had to fill out the English form, and could additionally fill out the Spanish form, if their child was exposed to the two languages.

ASQ-SE

For the home-visiting experiment and Longitudinal study, we use the Ages and Stages Questionnaire: Social Emotional (ASQ-SE) (Squires et al. (2002)) to measure social-emotional skills. The questionnaire is age-specific and completed by parents. They answer a series of questions such as "Does your child like to be hugged or cuddled?", "Does your child hurt himself on purpose?" on a three-point scale from "Never" to "Most of the time", with the possibility to indicate whether the described behavior is a concern for them. It is a tool used by pediatricians to detect potential issues in the social-emotional development of the child, when the overall score falls below a given threshold.

PLS

In the Longitudinal study (Table 2), the Preschool Language Scale (PLS) Fifth Edition was used to assess the language development of the children. The assessment is play-based, it is done by specifically trained assessors and covers a large range of linguistic skills that are specific to the age of the child (Zimmerman et al. (2013)).

Woodcock-Muñoz, ROWPVT and PPVT

In the home-visiting program, the Verbal Analogies test and the Picture Vocabulary test of the Woodcock-Muñoz Language survey were used to measure the expressive vocabulary of children (Woodcock et al. (2005)). The analogy test consists of a direct assessment in which a specifically trained assessor asks the child to complete a series of sentences (e.g., "A hairbrush is for hair and a toothbrush is for ..."). For the picture vocabulary test, the trained assessor points to a picture and asks the child to name the object (e.,g., a ball, a house, an apple).

To measure receptive vocabulary, the Receptive One-Word Picture Vocabulary Test (ROWPVT) Fourth Edition was used (Brownell (2010)). It is a direct assessment done by a specifically trained assessor who asks the child to point to specific objects or actions on an image that (s)he shows to the child. All scores were transformed in z-scores and aggregated in an overall vocabulary index (disaggregated results are presented in Table A.4).

The Longitudinal study (Table 2) uses another version of the picture vocabulary test called the Peabody Picture Vocabulary Test (PPVT) (Dunn and Dunn (2007)). It is also a direct assessment of receptive vocabulary based on a series of images presented to the child.

Give-N task

To measure math skills, we use the Give N task, which consists of providing the child with a large bowl of small items and asking him or her to give a certain number of items (e.g., "Can you give me TWO bananas?").

D Econometric model

For all our outcomes of interest, we estimate the following intent-to-treat model of being encouraged to watch the videos in the newborn study or receive home visits in the second study, separately for each assessment a:

$$Y_i^a = \alpha^a + \beta^a Z_i + \epsilon_i^a \qquad (1)$$

where Z_i is a dummy variable indicating whether parent i was assigned to the treatment group rather than the control group, $a \in \{6,\ 7,\ 9,\ 12 \text{ or } 18 \text{ months}\}$ for the newborn study and $a \in \{30\text{-}36 \text{ or } 36\text{-}42 \text{ months}\}$ for the home-visiting study. The outcomes Y are all standardized with respect to the earliest time point they were measured. We use the full sample size when the earliest time point is pre-intervention and the control group sample only when the earliest time point is post-intervention. β should thus be interpreted as an effect size. We estimate the model separately for each assessment to allow the coefficients and unobserved shock to vary at each stage of development. In section F, we present two variations of the specification shown in equation (1): one with control variables that have been selected based on the method described in Belloni et al. (2014), and one that distinguishes control group parents who just received usual care from control group parents who were in the placebo group and watched a safety video for the newborn study. The results tables present p-values that are adjusted for multiple hypothesis tested. Those adjusted p-values are computed following the method described in List et al. (2019). The method consists of controlling for the risk of false positive at the level of a family of outcomes. We define one family as one type of outcome (belief, interaction or child outcome) at a given age. Following that rule, we have the following eleven families for the newborn study:

1. Beliefs about impact of parents' inputs at 6 months
2. Beliefs about impact of parents' inputs at 7 months
3. Beliefs about impact of parents' inputs at 12 months
4. Beliefs about impact of parents' inputs at 18 months
5. Interactions with child at 6 months + Interactions with parent at 6 months
6. Interactions with child at 9 months + Interactions with parent at 9 months

7. Interactions with child at 12 months + Interactions with parent at 12 months
8. Interactions with child at 18 months + Interactions with parent at 18 months
9. Vocabulary at 9 months
10. Vocabulary at 12 months
11. Vocabulary at 18 months

And the following six families for the home-visiting study:

1. Beliefs about impact of parents' inputs at 30-36 months
2. Beliefs about impact of parents' inputs at 36-42 months
3. Interactions with child at 30-36 months + Interactions with parent at 30-36 months
4. Interactions with child at 36-42 months + Interactions with parent at 36-42 months
5. Social-emotional skills at 30-36 months
6. Social-emotional skills + Vocabulary + Math skills at 36-42 months

Note that most our input and outcome measures correspond to indexes that aggregate several measurements (see Appendix C), which restricts (ex-ante) the number of hypotheses we test, hence the small size of our families and the limited (ex-post) adjustment we impose to the p-values. Non-aggregated results can be found in Appendix E.

E Additional impacts

Table A.3: Impacts of the Newborn program at different time points

	6 months		7/9 months		12 months		18 months	
	$\overline{T}-\overline{C}$	N	$\overline{T}-\overline{C}$	N	$\overline{T}-\overline{C}$	N	$\overline{T}-\overline{C}$	N
Parents' beliefs on:								
Impact of parents' inputs	0.82***	385	0.61***	284	0.61***	375	0.38***	323
	0.00^{pc}		0.00^{pc}		0.00^{pc}		0.00^{pc}	
	0.00^{ad}		0.00^{ad}		0.00^{ad}		0.00^{ad}	
Parental self-efficacy			-0.11	296	-0.03	377		
			0.56^{pc}		0.73^{pc}			
			0.56^{ad}		0.73^{ad}			
Intelligence malleability					0.10	367	-0.01	315
					0.37^{pc}		0.94^{pc}	
					0.59^{ad}		0.94^{ad}	
Parents' inputs:								
Sensitivity to cues	-0.06	363	0.29**	341	0.19	374	-0.04	322

Continued on next page...

... table A.3 continued

	6 months		7/9 months		12 months		18 months	
	$\overline{T}-\overline{C}$	N	$\overline{T}-\overline{C}$	N	$\overline{T}-\overline{C}$	N	$\overline{T}-\overline{C}$	N
	0.60^{pc}		*0.01*^{pc}		*0.09*^{pc}		*0.73*^{pc}	
	0.60^{ad}		*0.05*^{ad}		*0.28*^{ad}		*0.97*^{ad}	
Response to distress	0.13	363	-0.06	341	0.16	374	0.05	322
	0.22^{pc}		*0.66*^{pc}		*0.25*^{pc}		*0.75*^{pc}	
	0.60^{ad}		*0.66*^{ad}		*0.44*^{ad}		*0.94*^{ad}	
SE growth fostering	0.09	363	0.19	341	0.12	374	0.02	322
	0.35^{pc}		*0.06*^{pc}		*0.19*^{pc}		*0.80*^{pc}	
	0.56^{ad}		*0.11*^{ad}		*0.46*^{ad}		*0.80*^{ad}	
Cognitive growth fostering	0.11	363	0.26*	341	-0.11	374	0.15	322
	0.30^{pc}		*0.04*^{pc}		*0.34*^{pc}		*0.17*^{pc}	
	0.63^{ad}		*0.10*^{ad}		*0.34*^{ad}		*0.48*^{ad}	
Child outcomes:								
Clarity of cues	-0.01	363	0.23*	341	-0.09	374	0.01	322
	0.92^{pc}		*0.03*^{pc}		*0.35*^{pc}		*0.93*^{pc}	
	0.92^{ad}		*0.06*^{ad}		*0.35*^{ad}		*0.93*^{ad}	
Responsiveness	-0.06	363	0.19*	341	-0.13	374	0.18	322
	0.53^{pc}		*0.07*^{pc}		*0.14*^{pc}		*0.09*^{pc}	
	0.76^{ad}		*0.07*^{ad}		*0.24*^{ad}		*0.16*^{ad}	
Words understood			0.19	319	0.22	369		
			0.15^{pc}		*0.15*^{pc}			
			0.26^{ad}		*0.25*^{ad}			
Words understood+said			-0.00	319	-0.00	369		
			1.00^{pc}		*0.98*^{pc}			
			1.00^{ad}		*0.98*^{ad}			
Words said							-0.10	321
							0.39^{pc}	
							0.60^{ad}	
Combines words							0.00	320
							0.94^{pc}	
							0.94^{ad}	

Note: $\overline{T}-\overline{C}$ gives the effect size of the intervention at 6, 7/9, 12 or 18 months. Below in italics are the per comparison (multiplicity-unadjusted) bootstrap p-value (pc), and the multiplicity-adjusted p-value (ad), both based on a two-sided Student's t test. The adjustment procedure follows List, Shaikh and Xu (2019). Families correspond to the set of four parental inputs, the set of two child's behaviors and the set of two vocabulary measures at each time point. N gives the number of observations used in the regression. In the 7/9 months column, beliefs are measured at 7 months, and behaviors are measured at 9 months. * for multiplicity-adjusted p-value<0.1, ** for multiplicity-adjusted p-value<0.05, *** for multiplicity-adjusted p-value<0.01.

Table A.4: Impacts of the home-visiting (HV) program at different time points

	30-36 months		36-42 months	
	$\overline{T}-\overline{C}$	N	$\overline{T}-\overline{C}$	N
Parents' beliefs on:				
Impact of parents' inputs	1.46***	69	1.25***	68
	0.00^{pc}		*0.00*^{pc}	
	0.00^{adj}		*0.00*^{adj}	
Parental self-efficacy	0.25	59		

Continued on next page...

... table A.4 continued

	30-36 months		36-42 months	
	$\overline{T}-\overline{C}$	N	$\overline{T}-\overline{C}$	N
	0.17^{pc}			
	0.17^{adj}			
Intelligence malleability	1.05***	69		
	0.00^{pc}			
	0.00^{adj}			
Child outcomes:				
WM Verbal Analogies English			-0.01	61
			0.97^{pc}	
			0.97^{adj}	
WM Picture Vocabulary English			0.07	61
			0.77^{pc}	
			0.93^{adj}	
ROWPVT English			0.15	67
			0.52^{pc}	
			0.80^{adj}	
WM Verbal Analogies Spanish			0.44	61
			0.14^{pc}	
			0.38^{adj}	
WM Picture Vocabulary Spanish			0.67**	61
			0.01^{pc}	
			0.04^{adj}	
ROWPVT Spanish			0.49	67
			0.05^{pc}	
			0.20^{adj}	

Note: $\overline{T}-\overline{C}$ gives the effect size of the intervention. In italics are the per comparison (multiplicity-unadjusted) bootstrap p-value (pc), and the multiplicity-adjusted p-value (ad), both based on a two-sided Student's t test. The adjustment procedure follows List, Shaikh and Xu (2019). Families correspond to the set of belief measures used at each time point and to the set of six vocabulary measures for the 36-42-month assessment. N gives the number of observations used in the regression. * for multiplicity-adjusted p-value<0.1, ** for multiplicity-adjusted p-value<0.05, *** for multiplicity-adjusted p-value<0.01.

F Robustness analysis

This section presents the average treatment effects (ATEs) discussed in the paper under two alternative specifications, with control variables and separating the two control subgroups for the Newborn experiment.

F.1 ATEs with controls

The choice of the control variables to add was dictated by two concerns: controlling for differences between the control and treatment groups that can occur due to small sample size and attrition, and increasing the precision of the estimates. To address both concerns, we rely on the method described in Belloni et al. (2014). The method consists in selecting the variables that best predict the treatment status with a LASSO regression (first concern), and then selecting the variables that best predict the outcome, with a LASSO regression as well (second concern). In the final regression (equation (1)), we include the union of the two sets of selected variables as controls. The first LASSO regression was run at each assessment point to account for possible differences between the treatment and control groups that would arise with attrition.

Table A.5 below shows that impacts on parental beliefs, inputs, and child outcomes are very similar to those presented in the paper (Table 1) when we add control variables.

Table A.5: Impacts of the two experiments at different time points with controls

	Newborn program				HV program	
	6m	7/9m	12m	18m	30-36m	36-42m
Parents' beliefs on:						
Impact of parents' inputs	0.69***	0.53***	0.46***	0.36***	1.31***	1.12***
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
	383	282	374	321	68	67
Parents' inputs:						
Interactions w/ child°	0.13	0.25**	0.12	0.08	0.86***	0.44
	0.24	*0.04*	*0.28*	*0.49*	*0.01*	*0.17*
	362	340	374	321	61	60
Child's outcomes:						
Interactions w/ parent	-0.05	0.21**	-0.11	0.12	0.66***	0.29
	0.59	*0.05*	*0.21*	*0.21*	*0.00*	*0.33*
	362	340	374	321	61	60
Vocabulary°		0.10	0.07	-0.02		0.35**
		0.31	*0.58*	*0.77*		*0.02*
		318	369	318		60
Math skills						0.48*
						0.05
						66
Social-emotional skills					0.49**	0.28**
					0.03	*0.04*
					69	67

Note: Differences between treatment and control group means are shown first for the two experiments at different time points. Below in italics are the corresponding p-values based on two-sided Student's t tests without adjustment for multiple comparisons. N gives the number of observations. In the 7/9 months column, beliefs are measured at 7 months, and behaviors are measured at 9 months. ° signals that the measurement tools used in the two experiments are different (see Section 4 of Supplementary Information). In the 7/9 months column, beliefs are measured at 7 months, and interactions are measured at 9 months. * for p-value<0.1, ** for p-value<0.05, *** for p-value<0.01.

F.2 ATEs distinguishing the two control groups (Newborn program)

In the Newborn program, the randomization was done in two stages. We first randomized parents between treatment and control groups. Then within the control group, half of the parents were randomized to watch a safety video (the topics covered were safety of the baby in the car, in case of a fall, fire, burn, choke, or poisoning), and the other half would not get any intervention and simply receive usual care at the clinic. The protocol followed for the safety videos was the same as the one we followed for the treatment videos. The rationale for having a placebo control group watching videos on a different topic was to test whether changes in parents' outcomes are simply driven by the attention they receive as part of the experiment or by the actual content of the video.

Table 1 in the paper merges the two control subgroups. Table A.6 below compares the treatment group to the control group who did not receive any video, and to the control group who received the safety video.[3] It shows that impacts on beliefs are similar when we distinguish the two control groups, confirming that changes in parental beliefs presented in the paper are driven by the content of the language videos.

Results on parents' interactions with their child are similar when we separate the two control groups as well. However, the impacts on children's interactions tend to be slightly different when we distinguish between the two control groups. Table A.6 shows that children's behaviors are also positively affected in the safety control group at 9 months, so the treatment effects are higher when we compare the language treatment group to the usual care group only (0.4σ versus 0.2σ in Table 1).

Table A.6: Impacts of the Newborn program at different time points (separate control subgroups)

	6 months		7/9 months		12 months		18 months	
	$\overline{T}-\overline{C'}$	$\overline{S}-\overline{C'}$	$\overline{T}-\overline{C'}$	$\overline{S}-\overline{C'}$	$\overline{T}-\overline{C'}$	$\overline{S}-\overline{C'}$	$\overline{T}-\overline{C'}$	$\overline{S}-\overline{C'}$
Parents' beliefs on:								
Impact of parents'	0.85***	0.05	0.54***	-0.13	0.63***	0.03	0.37*	-0.02
inputs	0.00^{pc}	0.72^{pc}	0.00^{pc}	0.37^{pc}	0.00^{pc}	0.81^{pc}	0.02^{pc}	0.90^{pc}
	0.00^{ad}	0.72^{ad}	0.00^{ad}	0.69^{ad}	0.00^{ad}	0.96^{ad}	0.05^{ad}	0.90^{ad}
Parental			-0.05	0.12	0.06	0.18		
self-efficacy			0.85^{pc}	0.66^{pc}	0.62^{pc}	0.13^{pc}		
			0.85^{ad}	0.84^{ad}	0.94^{ad}	0.44^{ad}		
Intelligence					0.02	-0.15	-0.18	-0.33*
malleability					0.87^{pc}	0.31^{pc}	0.17^{pc}	0.04^{pc}
					0.87^{ad}	0.74^{ad}	0.31^{ad}	0.10^{ad}
Parents' inputs:								
Interactions	0.21	0.16	0.26	-0.02	0.13	0.06	0.05	-0.10
w/ child	0.11^{pc}	0.28^{pc}	0.09^{pc}	0.93^{pc}	0.32^{pc}	0.71^{pc}	0.76^{pc}	0.55^{pc}
	0.26^{ad}	0.47^{ad}	0.14^{ad}	0.93^{ad}	0.74^{ad}	0.71^{ad}	0.76^{ad}	0.76^{ad}
Child's outcomes:								
Interactions	0.11	0.31	0.40**	0.33*	-0.09	0.09	0.18	0.10
w/ parent	0.39^{pc}	0.03^{pc}	0.01^{pc}	0.03^{pc}	0.43^{pc}	0.48^{pc}	0.16^{pc}	0.49^{pc}
	0.39^{ad}	0.11^{ad}	0.02^{ad}	0.09^{ad}	0.78^{ad}	0.73^{ad}	0.46^{ad}	0.85^{ad}
Vocabulary			0.15	0.11	0.21	0.19	-0.03	0.02
			0.16^{pc}	0.41^{pc}	0.15^{pc}	0.35^{pc}	0.73^{pc}	0.81^{pc}
			0.28^{ad}	0.41^{ad}	0.26^{ad}	0.35^{ad}	0.92^{ad}	0.81^{ad}

Continued on next page...

[3]Note that the p-value adjustment follows the same families as described in section D, but the size of the families is multiplied by two here, as we have two control groups.

... table A.6 continued

	6 months		7/9 months		12 months		18 months	
	$\overline{T}-\overline{C'}$	$\overline{S}-\overline{C'}$	$\overline{T}-\overline{C'}$	$\overline{S}-\overline{C'}$	$\overline{T}-\overline{C'}$	$\overline{S}-\overline{C'}$	$\overline{T}-\overline{C'}$	$\overline{S}-\overline{C'}$

Note: $\overline{T}-\overline{C'}$ (resp. $\overline{S}-\overline{C'}$) gives the effect size of the intervention at 6, 7/9, 12 or 18 months for TMW (resp. Safety) video vs Usual care group. Below in italics are the per comparison (multiplicity-unadjusted) bootstrap p-value (pc), and the multiplicity-adjusted p-value (ad), both based on a two-sided Student's t test. The adjustment procedure follows List, Shaikh and Xu (2019). Families of outcomes for the adjustment are described in Section 5 of Supplementary Information. In the 7/9 months column, beliefs are measured at 7 months, and behaviors are measured at 9 months. * for multiplicity-adjusted p-value<0.1, ** for multiplicity-adjusted p-value<0.05, *** for multiplicity-adjusted p-value<0.01. The number of observations is the same as in Table 1 (for Parental self-efficacy and Intelligence malleability, please refer to Table A.4).

G Do changes in beliefs cause changes in parental investments?

This section explores further the relationship between the changes in parental beliefs about early skill formation and the changes in parental investments in children (see section 3.1 for a full presentation of our economic framework). To do so, we exploit the random variation induced in parents' beliefs by each intervention in an attempt to estimate the causal effect on parents' investments via 2SLS regressions.

In the first stage, we regress our measure of beliefs on the treatment status variable. In the second stage, we regress our measures of parental investments on the predicted beliefs recovered from the first stage. This technique yields an estimate of the effect of parents' beliefs on their investments that can be interpreted causally if the treatment status strongly predicts beliefs (which is verified in Table 1), and if the treatment has no direct impact on parents' investments (no impact other than through beliefs improvement). Although the second condition cannot be directly tested, we argue that since the interventions mainly consist of providing information about child development and parents' potential to affect it, and of teaching parents how to improve their parenting practices, it is plausible that one of the main channel through which the interventions affect parents' investments with their child is by changing their beliefs and knowledge of parenting, which is captured by our SPEAK survey (see Appendix C).[4]

Table A.7 presents the results at the same time points as Table 1. The first four columns of the table show the impact of beliefs measured at 6, 7, 12, and 18 months on parental inputs measured at respectively 6, 9, 12, and 18 months, using the Newborn data. The last two columns show those impacts at different ages, using the home-visiting data. At t=6 months, we fail to reject the hypothesis that changes in parental beliefs lead to changes in their investments (hypothesis $\frac{\partial I_t}{\partial \phi_t} = 0$ in our economic framework, section 3.1), which is consistent with the fact that we find no impact of the intervention on investments at that age (Table 1). However, the second row of the table reveals that improved knowledge of the impact of parental investments at 6 months

[4]This assumption is stronger in case of the home visiting program, we discuss potential other channels at the end of this section.

leads to improved parental investments at 9 months ($\frac{\partial I_t}{\partial \phi_{t-\ell}} > 0$, for t=9 months, ℓ= 3 months).[5] Similarly, the home-visiting data reveal that improvements in parents' beliefs at 30-36 months induce contemporaneous improvements in their investments, consistent with the treatment effect on parents' investments we find at that age in Table 1. Other impacts shown in the table all have a positive sign, but are too small to be significant given the number of observations we have. These results support the hypothesis that the impacts we find on parents' investments at 9 months in the Newborn experiment and at 30-36 months in the home-visiting experiment are driven by learning.

Table A.7: Impact of parents' beliefs about the impact of parental investments on their investments at different time points

	Newborn program				HV program	
Beliefs at:	6m	7m	12m	18m	30-36m	36-42m
Investments at:						
6m	0.16					
	0.21					
	363					
9m	0.35**	0.31				
	0.03	*0.20*				
	340	257				
12m	0.16	0.22	0.17			
	0.29	*0.36*	*0.38*			
	362	272	372			
18m	0.12	0.14	0.15	0.28		
	0.43	*0.61*	*0.53*	*0.36*		
	312	232	308	321		
30-36m					0.45*	
					0.06	
					61	
36-42m					0.33	0.37
					0.17	*0.18*
					59	60

Note: Each row gives the change in parents' investments (in standard deviations of the control group) that is associated with an increase of the beliefs score by one standard deviation (the beliefs score is instrumented with the treatment status variable). The second row gives the corresponding p-value in italics based on a two-sided Student's t test without adjustment for multiple comparisons. * for p-value<0.1, ** for p-value<0.05, *** for p-value<0.01. The third row gives the number of observations.

To take into account the possibility that the interventions also affect parental investments through other channels such as social pressure (by communicating norms about what "good parents" should do) or accountability (due to the monitoring of their behavior by the research team), we re-did the estimations controlling for parents' perceptions of the pressure they face (measured by the TOol to measure Parenting Self-Efficacy (TOPSE) - Pressure subscale). Results remain similar in the Newborn program but shrink and become statistically non-significant in the HV program,

[5]If we decompose the measure of parental investments, we find that the overall impact is driven by improvement in parents' sensitivity to cues, social-emotional and cognitive growth fostering (results available upon request).

suggesting that the exclusion condition described above might be less plausible in the latter case (table available upon request).

H Experimenter demand effect

In our studies, experimenter demand effects could happen in two ways. First, the intervention that treated parents receive gives them cues about what the experimenter considers to be desirable attitudes, and the higher scores we observe in self-reported surveys could merely be due to parents learning about what we hope to find. While this could be true at the six-month assessment for the Newborn program, for example, which takes place right after parents watch the last video, it is less likely to persist in the following assessments. Yet, one year after the end of the intervention, at the eighteen-month assessments, treated parents still score 0.4σ higher than control parents. It should also be noted that in both experiments, it was made clear that answers to surveys would remain anonymous and be only analyzed in aggregation, which likely mitigated incentives for parents to try to please the experimenter.

Another way in which experimenter demand effects could happen is through the surveys. Asking parents about their attitudes with their child could indeed induce them to search for information about the topics covered in the survey and, if that searching behavior is asymmetric between treated and control parents (if control parents search more, for instance, because they do not receive information about child language development while the survey elicits their beliefs about it), our treatment effects would be biased (downward if control parents exhibit compensating behaviors). Nonetheless, Figure A.3 below shows that the evolution of beliefs in the control group between the baseline and following assessments is rather stable across both field experiments, suggesting that control parents do not update their beliefs over the period of study. Furthermore, this fear is attenuated when we consider the link between beliefs and actions.

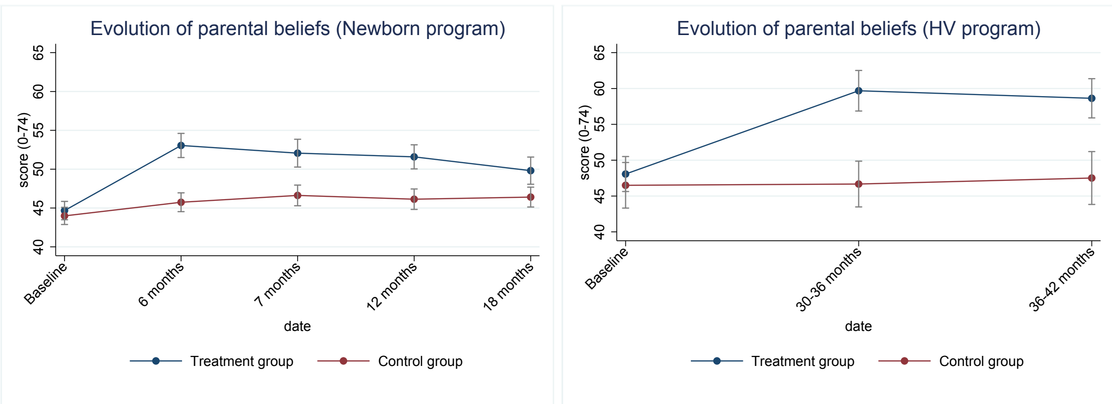


Figure A.3: Evolution of parents' beliefs about how parental investments affect child development at different time points