

Parental Beliefs, Investments, and Child Development

Evidence from a Large-Scale Experiment

Pedro Carneiro
Emanuela Galasso
Italo López García
Paula Bedregal
Miguel Cordero



WORLD BANK GROUP

Development Economics
Development Research Group
February 2019

Abstract

This paper experimentally evaluates a large-scale and low-cost parenting program targeting poor families in Chile. Households in 162 public health centers were randomly assigned to three groups: a control group, a second group that was offered eight weekly group parenting sessions, and a third group that was offered the same eight group sessions plus two sessions of guided interactions between parents and children focused on responsive play and dialogic reading. Three years after the end of the intervention, the receptive vocabulary and the socio-emotional development of children of families participating in either of the treatment arms improved (by 0.43 and 0.54 standard deviation,

respectively) relative to children of nonparticipating families. There were no statistically detectable impacts on other types of skills. The treatments also led to improvements in home environments and parenting behaviors of comparable magnitudes, which far outlasted the short duration of the intervention. A simple mediation analysis suggests that up to 13 percent of treatment impacts on language, and up to 36 percent of impacts on child socio-emotional development, can be attributed to changes in the home environment, as well as in nurturing and discipline parenting behaviors.

This paper is a product of the Development Research Group, Development Economics. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/research>. The authors may be contacted at egalasso@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Parental Beliefs, Investments, and Child Development: Evidence from a Large-Scale Experiment¹

Pedro Carneiro
University College London,
IFS and CeMMAF

Emanuela Galasso
World Bank

Italo López García
RAND Corporation

Paula Bedregal
Pontificia Universidad Católica, Chile

Miguel Cordero
University of Bristol

Keywords: parenting, early childhood development

¹ We thank participants in several seminars and conferences for comments. This project was funded by the World Bank Research Budget and the Strategic Impact Evaluation Fund. Pedro Carneiro gratefully acknowledges the support of the ESRC for CEMMAP (ES/P008909/1) and the European Research Council through grant ERC-2015-CoG-692349. Miguel Cordero acknowledges to the Scholarship Formation of the Advanced Human Capital Programme, Becas Chile. A special thank goes to Lucia Vergara, Cecilia Moraga and Felipe Arriet at the Ministry of Health for their support of the evaluation, to Veronica Silva for her help in setting up the evaluation, and to Veronica Mingo and the staff of *Juguemos con Nuestros Hijos* for the adaptation of the curriculum for the intensive arm of the evaluation. We are grateful to Ruben Poblete Cazenave and Nicolas Libuy Rios for able research assistance. Finally, our greatest thanks go to all parents and children who participated in this study and to the local facilitators and their primary care teams who helped to make this study happen. All errors and omissions are own.

1 Introduction

A large body of research shows that there are high potential returns to early childhood programs, especially when they are targeted to disadvantaged families (Heckman 2006; Heckman and Masterov 2007). However, not all early childhood interventions are equally successful. Early childhood interventions have a larger chance of producing long lasting impacts when they also lead to permanent changes to parental behaviors, which outlast the duration of the program. When this occurs, and the quality of the environments in which children grow up is changed, we can see as a result sustained changes in child health, cognitive and socio-emotional skills. Examples of interventions which have produced long-term impacts on child development and where home environments played a key role in these changes are Abecedarian (Campbell and Ramey 1994) or the Jamaica Study (Gertler et al. 2014; Grantham-McGregor et al. 1991).

This paper documents the medium-term experimental impacts on home environments and child development of a low-cost and short duration group parenting program, which was delivered at scale in Chile. This program aims to affect child development by promoting sustained changes in parenting behaviors. The program is the Chilean adaptation of the Canadian program “Nobody is Perfect”, which has operated in the Canadian public health system for more than 30 years. The adapted version of the program is called “Nadie es Perfecto” (NEP hereafter).

NEP aims to improve the quantity and quality of parental investments in children by providing parents with the information, motivation and self-confidence to implement positive parenting behaviors. This program is delivered by trained local public health workers who organize weekly group sessions with caregivers of children of 0-5 years old, using an experiential learning model through group discussions. We show that, despite its short duration and intensity, the program produces impacts on home environments and children’s language development, lasting well beyond the conclusion of the program.

Our data come from a large-scale field experiment with two treatment arms and a control group. Parents in the control group have access to regular health preventive services offered by the primary health care system.² Parents in the first treatment arm (NEP-Basico, or NEP-B) are invited to participate in a group-parenting program where only caregivers are allowed to attend the meetings (no children), consisting of 6 to 8 weekly group sessions with 6-12 caregivers. The second arm (NEP-Intensivo, or NEP-I) complements the caregiver-only sessions with two additional structured sessions where children participate together with their parents, and which are focused on the importance of language and play. The target population are parents with children aged 0 to 5 who are enrolled in the public health system, especially those who are particularly vulnerable, such as adolescents, single parents, and geographically or socially isolated households.

The duration of the intervention is short, and our focus is on its medium-term impacts, measured long after program completion. A full program cycle of 6-8 sessions lasts only for two to three months. We administered a baseline survey in 2011, right before parents in the treatment groups received the invitation to participate in the program, and re-surveyed the same parents and respective children in 2014, three years after the conclusion of these households’ participation in

²The national child health care program includes services such as developmental screenings, vaccinations, nutritional supplementation and regular check-ups until the child is 6 years old.

the intervention.

We find that children whose parents participated in NEP experienced an increase of 0.43 standard deviation (SD) in a measure of cognitive development (receptive vocabulary), and an increase of 0.54 SD in a measure of personal-social development, relative to children in the control group. There were however no detectable medium-term impacts of NEP on executive function, or on internalizing and externalizing behaviors of children (point estimates are much closer to zero, sometimes negative, and never statistically significant). It is possible that positive shifts in these outcomes occurred shortly after the program ended and did not last, although we cannot verify this hypothesis.

The impact on language is mirrored by increases of similar or larger magnitude in indicators of the quality of the home environment, nurturing practices and a reduction in the report of harsh disciplinary strategies. Importantly, these changes are supported by significant improvements of similar magnitude in perceived parenting self-efficacy, in parental beliefs about their role in child development, as well as by perceived social support from friends and the community. All these results point to sustained changes in the parental self-efficacy, beliefs and in the parent-child interaction as key pathways for the sustained changes in cognitive and socio-emotional development.

These are remarkable results. They show how a parenting program which has low cost and is easy to implement at a national scale can produce long lasting changes in the lives of poor children and their families. It provides an interesting model for other countries, especially those struggling to provide valuable early childhood services due to severe lack of resources. One needs to consider, however, that Chile's welfare system is well organized and staffed with a skilled and motivated workforce, which may have played a role in their success in implementing NEP.

Despite growing evidence on the effectiveness of parenting programs evaluated in small-scale trials, key questions remain relating to how to make these programs scalable, and how to sustain their impacts long-term. Individual home visits provide an opportunity to tailor activities to individual circumstances and personal barriers to behavioral change, but they are expensive to implement at scale.³ Group parenting programs are less costly to implement at scale and favor the creating of support networks, but they offer less individualized attention and guided practice. The evidence of these programs, which is the type of program we examine here, is somewhat mixed.⁴ Group-based interventions have been shown to shift parental knowledge about child development

³ A home visiting model in Jamaica that provides individualized support to parents to help them improve their child development found large effects on children's developmental outcomes, sustained until adulthood (S. M. Grantham-McGregor et al. 1991, Gertler et al. 2014). Cross country replication of the Jamaican home visiting curriculum to other settings (such as Bangladesh, India and Colombia) has so far found impacts on child development in the short-term (O. P. Attanasio et al. 2014; Hamadani et al. 2006; Nahar et al. 2012; Vazir et al. 2013).

⁴ The systematic evidence in the LANCET series on Early Childhood Development suggests that interventions that directly engage the child in play experiences and guide caregivers in providing stimulating interactions, as well as give parents the opportunity to practice with their children and receive feedback, can be effective in remediating early disparities in child development (Black et al. 2017; Britto et al. 2016; Engle et al. 2007, 2011). A recent systematic meta-analysis, on the other hand, finds mean effect sizes in children's language and cognitive outcomes of 0.32 SD for home-based studies, and 0.59 SD for group visit studies (Aboud and Yousafzai 2015).

and promote positive parenting behaviors, although not systematically improving child development outcomes (Aboud and Akhter 2011, Singla, Kumbakumba, and Aboud 2015, Yousafzai et al. 2014). Group meetings do not offer as much personalized attention, but provide social support and peer-to-peer learning to support changes, and are potentially more cost-effective (Aboud and Yousafzai 2015). Moreover, there is no evidence of a scalable group-based program with sustained impacts in child outcomes and parental behaviors.

We contribute to this literature by experimentally studying a parenting intervention implemented at scale within the health system of Chile, relying on existing infrastructure and human resources. This is in sharp contrast with most experimental evaluations in this field, which focus on relatively small pilot programs (Richter et al. 2017). There are few notable exceptions. In developed settings, programs such as the Nurse Family Partnership program (Heckman et al. 2017) or the Preparing for Life program in Ireland (Doyle et al. 2013) rely on high frequency contact (weekly or bi-weekly contacts from pregnancy to two or to five years) with highly trained and professional home visitors. The NFP program involves weekly or bi-monthly visits to low-income mothers from pregnancy to 2 years of age. The PFL program, modeled after the NFP, included bi-monthly home visits from pregnancy to age 5, supplemented by group parenting classes similar to the NEP. In LMIC settings, the Lady Health Worker program in Pakistan is a mixed model of nutrition and early stimulation interventions combining a few group sessions with a more intensive and extended period of individual home visits that was implemented at scale and has been proven effective (Gowani et al. 2014). The second example comes from the recent evaluation of the FAMI program in Colombia, where a structured curriculum modeled after the Jamaica experience was added to an existing program delivering weekly group activities and a monthly home visit (Attanasio et al. 2018).

The NEP program stands out in contrast to these programs implemented at scale in three important ways. First, NEP is delivered in groups, which makes it less costly to implement at scale. Based on calculations from the Ministry of Health, the cost per child attended by NEP is roughly 10 times cheaper than a home visit.⁵ Second, its intensity and duration. The program we are studying here is significantly shorter (6-8 sessions) with a much lower frequency and intensity of contact, though we try to bring a dimension of intensity/duration by adding an experimental arm that extends the basic design with two additional sessions with both parents and children. And third, an important unique feature of the intervention is its focus on the cycle of experiential learning. Parents draw from their own experiences, learn from each other and discuss main challenges they are facing and share strategies. NEP relies on a semi-structured curriculum that fosters parental competence by tailoring the intervention to the group interests and needs, rather than following a given sequence of planned activities.

We study medium-term program effects three years after the program ended. With two measurements (before and three years after the program), we cannot disentangle how much of the effects are a result of cumulative early gains or a result of continued improvements in parental investments. Our mediation analysis focuses on parental beliefs, as well as investments in the cognitive and socio-emotional home environment as potential mechanisms for the sustained gains

⁵ In 2011, the cost per hour and family attended of a standard home visit was US\$7.33, while the cost of NEP Basico was US\$0.63 and the cost of NEP Intensivo was US\$0.85.

in child development outcomes in the medium run. The intervention permanently fostered improved parenting behaviors, parental beliefs about their role in child development, perceived social support and parental self-efficacy. Our mediation analysis suggests that changes in the measures of parental behaviors and perceptions can account to as much as 13% of the impacts on child language and up to 36% of the impacts on personal and socio-emotional development.

The paper proceeds as follows: Section 2 presents the study design, Section 3 describes the data, Section 4 shows the empirical strategy, Section 5 discusses our findings, and Section 6 concludes.

2 Study Design

2.1 The intervention

NEP is a parenting intervention operating in the context of a broader early childhood policy platform called Chile Crece Contigo (ChCC). The intervention was adapted from the Nobody's Perfect program in Canada, a long-running group parenting intervention implemented within the public health system in Canada. NEP relies on a semi-structured curriculum that promotes knowledge about child development, parental self-care, positive parenting skills in caregivers, and the use of non-violent disciplinary strategies, helping caregivers to foster a nurturing home environment.

NEP targets parents with children aged 0 to 5 who are enrolled in the public health system. Potential participants are offered participation in the program during the regular health check-ups, home visits or immunization visits. The intervention can be applied to all parents who are interested in improving their parental skills, but it is more directly targeted to caregivers who are particularly vulnerable, such as adolescents, single parents, and geographically or socially isolated households. Parents in these groups, and other parents who are in need of this type of intervention, can be identified by the health care provider (doctor or nurse) with whom they interact frequently. Households at very high risk (children with severe child developmental delays or disabilities, or high-risk parents with psychiatric problems or intra-household violence) are not considered eligible for NEP and are instead referred to services with more intensive engagements at the local level.

The standard program (which we call NEP-Basic, or NEP-B) includes 6 to 8 weekly group sessions with 6-12 caregivers, facilitated by a trained moderator, and based on a curriculum that promotes positive parenting skills to improve cognitive stimulation, to manage child behavior with positive disciplinary strategies, and to improve their parental self-esteem. Each session lasts approximately two hours. An intensive version (NEP-Intensive, or NEP-I) was developed as part of the study as an additional evaluation arm. It adds to NEP-B two practical sessions with children in order to give caregivers the opportunity to interact with their child in a monitored environment and thereby receive more personalized feedback on their practices.

There are several features that distinguish NEP-B from other group interventions, and which are worth highlighting. The first key innovation of the approach lies in a semi-structured curriculum that fosters parental competence by tailoring the intervention to the group's interests and needs. This flexibility is what is novel and important, allowing parents to choose the specific topics for each

session (organized along physical development, cognitive development, behavior, safety and parental self-care).⁶

NEP is based on a model of experiential learning designed for adults,⁷ which aims to promote an active engagement in the group discussion and agency. Parents draw from their own experiences, learn from each other and discuss main challenges they are facing and share strategies. This model also fosters the creation of deep relationships between parents participating in the same group, which could potentially spillover to their lives outside the group, since all parents live within a relatively small area, which is served by the health center they attend.

The premise of the intervention is that to translate knowledge and beliefs into real behavioral change, participants need not only to increase their knowledge about the optimal practices, but also to emotionally connect to the way themes are discussed with other parents facing similar problems. Parental behavioral change also requires an improvement in parental self-image, as well as the perception of support by the network formed with shared norms of positive parenting practices (Kagitcibasi et al. 2009).

A second distinctive feature of NEP is that it combines highly qualified staff and high-quality training, which enable this program to be less susceptible to suffering from the problem of sustaining quality of service delivery at scale (Davis, Guryan, Hallberg, & Ludwig, 2017). NEP is delivered by facilitators who are local professional staff in the health centers (such as nurses, psychologists and social workers), who have frequent interaction and knowledge of the target population participating in the program given their close interaction through the health system. In addition, facilitators are trained on the NEP methodology by a set of master trainers, certified by the Canadian Nobody's Perfect Program. The focus of their training is on active listening skills, and on facilitating group dynamics with flexibility, without forcing themes or lecturing parents.

⁶The topics are covered in five books, which are distributed to participants:

1. Physical development, including topics such as physical growth, health, nutrition and early detection of common illnesses in early years.
2. Mental development, including topics such as cognitive and emotional development, the role of playing and how to stimulate a child according to their age.
3. Behavior, designed as a guide on common behavioral problems and their effective management and resolution using positive disciplinary strategies.
4. Safety and prevention, designed to identify, prevent and manage common risks and accidents at home, including first aid training.
5. Parental and caregiver's self-care, involving activities to improve parental self-image, self-help in the parenting task, the prevention of domestic violence and the promotion of healthy habits strategies for adults.

Both caregivers and facilitators are provided with additional materials (stickers with emergency phone numbers on them, promotional posters of NEP for parents, audiovisual and board games for facilitators).

⁷The training workshop looks at introducing facilitators to the model, learning the goals of the program, and how to use the criteria to select participants who would get benefit from the interventions. The main goal is that facilitators learn how to conduct a parenting course from the beginning to end; using a participant-centered method, implementing approaches for adult education, and following the Experience Learning Cycle, a well-established framework to understand how adults can learn (Kolb 2014).

The intensive version of the program (NEP-I) was not part of the set of services originally offered by the Ministry of Health and was developed especially for this study. In spite of that, it was adopted by the Ministry of Health staff administrating NEP and also delivered at scale during the evaluation period. NEP-I was a collaborative effort between the Ministry of Health and a team of child development experts at Pontificia Universidad Catolica (working on the program *Juguemos con Nuestros Hijos*). It adds to the standard group intervention two practical sessions, where caregivers are given the opportunity to interact with their child and receive more personalized feedback. The rationale for the intensive version is to test the value added of offering opportunities for practical demonstration and skill building, which has been shown elsewhere to be associated with effectiveness in parenting interventions (Engle et al. 2007). The two added sessions focus on the importance of age-appropriate responsive play (reading children's cues and providing scaffolding, through practice and discussion videos on sensitive play parent-child interactions) and on the importance of language and reading (through dialogic reading).

NEP-B has been fully scaled-up at the national level and is potentially highly cost-effective. It uses the infrastructure and human resources already existing in the health network with no further monetary and organization costs beyond training and material printing. The costs of NEP-B per family attended are only 10% of the costs of home visits. The more intensive version of the program being tested (NEP-I) costs 30% more than the standard version (NEP-B) per family attended.

2.2 The evaluation design

NEP was implemented across Chile. Therefore, our study is based on a representative sample of health clinics located in both urban and rural areas all over the country. The sample was stratified by type of clinic, which included family health centers, general health centers, and small hospitals (this stratification was motivated by the idea that different infrastructure and human resources across types of health centers may play an important role in the delivery of the program).

Within each clinic, a sample of 18 families was randomly drawn from a potential wait-list of participants formed by facilitators (which usually contains between 45 and 60 potential participants per center). Potentially eligible families⁸ were identified during regular health visits to the center and added to the wait-list just prior to the administration of the baseline survey.

⁸ The recruiting procedure is as follows. Facilitators construct and review the waitlists checking the clinical records of each family, with the purpose of selecting those families that satisfied both the inclusion and exclusion criteria of eligibility for NEP. Families in crisis such as detected domestic violence, severe mental health problems, or child developmental delays that require clinical attention are excluded from the group sessions. Health professionals at the health center derive families in these cases to individualized attention. Once identified, parents of eligible families were invited to participate through home visits or through direct recommendation made by a health professional. Eligible families were enrolled after an interview with an NEP facilitator, where they were informed about this study's intentions and about the randomized process of assignment to groups. Also, they were given the chance read the informed consent clause (or received an assisted reading of the same when they declared difficulties to read, poor reading skills or illiteracy). After accepting and signing the clause, they were included in a general database and the facilitators proceeded to randomly assign them to one of the groups. Parents who refused to participate continued receiving their usual health care offered at their usual health center and were encouraged to reconsider their entry for a future wave of NEP.

The 18 families selected to be part of the study were then randomly assigned to three groups: 1/3 was invited to participate in NEP-B, 1/3 was invited to participate in NEP-I, and the remaining 1/3 of families was assigned to the control group. The control group remained on a waiting list up until the endline survey was conducted, at which point they became eligible to participate in NEP. Families in the control group receive no NEP benefits, but they continued to receive their usual health care at the health center, which included non-structured talks with the parents and regular control visits to children. Treatment families were free to accept or not the invitation to participate in NEP. We discuss the extent to which they took up this invitation, and the consequences for our estimates.

The final sample includes 162 health clinics stratified by type of health center, 324 facilitators (162 for the basic NEP and 162 for the enhanced NEP-I), and 18 households per health center (6 treatment NEP-B + 6 NEP-I + 6 control), which resulted in a total sample size of 2,916 caregivers and 3,597 children evaluated at baseline.

2.3 Measurements

There are two survey waves used in this study: a baseline survey which occurred before the intervention took place, administered in June-September 2011; and an endline survey administered in July-October 2014, almost three years after the end of the group sessions for the sample of households participating in this study. The 6 to 8 week NEP program affecting these households and their children occurred in slightly different periods in each participating clinic, and they all occurred between October 2011 (start date for the first NEP group in the study) and March 2012 (end date for the last NEP group in the study). These surveys cover different dimensions of caregiver characteristics and behaviors as well as child outcomes, which we now describe.

2.3.1 Parental beliefs, attitudes and expectations

A set of variables we consider measure different dimensions of parental beliefs, attitudes and expectations. At least one-third of the sessions in NEP aims to promote participants' self-care and self-image as parents. This dimension of parental, perceptions, related to parental self-efficacy, is grounded in social cognitive theory (Bandura 1986, 1995)). In order to measure this concept, we use the Parenting Sense of Competence Scale (Ohan, Leung, and Johnston 2000), a 17-item scale that evaluates parental confidence in their capacity to overcome daily child-rearing tasks. A complementary instrument captures how parents perceive that their behavior has any impact on child outcomes. To this end, we adopt a subscale of the Parental Cognitions and Conduct Toward the Infant Scale (PACOTIS) (Boivin et al. 2005), a 5-item Likert scale to assess the perceived parental impact of their behavior on the developing child. We dichotomized the items and constructed a perceived impact indicator adding all items.

Social support has been signaled as an important mediator of change in group-based health interventions (Briscoe and Aboud 2012). To measure perceived social support by parents we used a short version of the Social Provision Scale (Cutrona and Troutman 1986), with subscales for perceived support from the family, from friends or the community, and from significant others.

To capture parental beliefs about how to raise children, in particular ideas about structure and warmth in child-rearing tasks, we adapt the Ideas About Parenting (IAP) questionnaire (Heming,

Cowan, and Cowan 1990). This scale can be used to characterize parenting in terms of authoritarian, authoritative, and permissive (see Baumrind (1968) and Maccoby et al. 1983).

Finally, we measure expectations about the benefits of a better home environment and quality time spent with children by adapting a scale to elicit these beliefs developed by Cunha, Elo, and Culhane (2013). This instrument asks parents about the age at which they would expect children to achieve developmental milestones in language and socio-emotional development, under different scenarios concerning home environments and parental investments. Our adaptation confronts parents with two scenarios: a high-investment home, in which parents spend quality time cognitively stimulating their children and do not use harsh disciplinary strategies to manage their behaviors; and a low-investment home, in which parents do not cognitively stimulate their children often and they rely on harsh discipline. Our hypothesis is that parents in the treatment group may report earlier ages in the high-investment home than those in the control group, because as a result of NEP they are more aware of the impact of home investments on child development.

2.3.2 Parental investments in children

To measure parenting behaviors and home environments we combine self-reported and directly observed variables. We use two sub-scales of the Parent Behavior Checklist (Fox 1994), where parents were asked to indicate how frequently they engaged in different activities with their child over the past couple of weeks. A first sub-scale measures Nurturing practices, associated to positive parental socio-emotional interactions with the child. A second sub-scale measures Discipline practices, a mixture of positive and harsh disciplinary practices (an exploratory factor analysis of the Discipline subscale confirmed that these are indeed capture two separate underlying constructs).

In addition, in the baseline survey we administered the Family Care Indicators (FCI) (Hamadani et al. 2010), which measures the frequency of learning and play activities with children, as well as the amount and variety of play and learning materials available at home. In the endline survey we used a revised version of the FCI, with additional self-report and observational items from the HOME-SF (Bradley and Caldwell 1984), enabling us to expand the Family Care Indicators for an older age group. Using these items (which are highly correlated to each other), we construct a latent factor index of cognitive stimulation using all the items using principal components analysis.

2.3.3 Child development outcomes

We consider three developmental domains potentially affected by the intervention: language, executive function, and socio-emotional development. Different test instruments available at baseline reached an age limit and endline, and were complemented by measurements that could cover a wider age range.

Language: At baseline we measured both receptive and expressive language for children from 0 to 71 months using the Spanish version of the Preschool Language Scale (PLS-4). However, because a large proportion of children at endline were older than 71 months and could not be administered the PLSIV, in the endline survey we applied the “Test de Vocabulario en Imágenes” (TEVI-R), a direct assessment for receptive vocabulary that has been adapted from the Peabody PPVT and normed for the Chilean context and was administered to children 36 months of age and

older (Echeverria, Herrera, and Segure 2002).⁹ We use standardized scores based on the suggested normalizations from the publishers of the test, as well as standardized scores based on the estimation of our own latent construct for receptive language using IRT methods. Results are similar regardless of the method used to construct these scores.

Executive function: These are the cognitive aspects of self-regulation (Blair and Razza 2007) and sometimes defined as working memory, inhibitory control, attention, and cognitive flexibility, which have been shown to be important predictors for children's social and academic development. We applied both at baseline and endline the Dimensional Change Card Sort (DCCS) task (Zelazo 2006), which is appropriate for longitudinal uses starting from age 2½ until adulthood. In the standard version of the test, children are asked to order a series of cards according to one dimension (for example, the color), and then according to another dimension (for example, the shape). The test requires holding two pieces of information in mind and at the same time inhibiting a dominant tendency when the task is switched. This is primarily a test of cognitive flexibility. At endline, we also administered a Leiter-R scale to measure the capacity to sustain attention.

Socio-Emotional Development: Approximately one-third of the group discussions in NEP-B were devoted to behavioral issues in children. This reflected the interests of program participants, who were looking for practical tools to address their children's behaviors. We use two measures to capture the range of behavioral problems (maladaptive behavior) as well as the positive socio-emotional development (adaptive behavior) as reported by the primary caregiver. We administered the Achenbach Child Behavior Checklist (CLBC, Achenbach and Ruffle 2000), which captures internalizing and externalizing behavioral problems for children aged 1½ years and older. In order to measure positive dimensions of how the child establish interpersonal relationships, we used the Battelle Developmental Inventory Screening Test (BDIST II) Personal- Social Scale. (Ringwalt 2008) We focused on the three subscales that capture three dimensions of the socio-emotional development of children: interaction with adults, interactions with peers, and the self-concept and social role. The first two subscales of BDIST II are available for children up to 71 months (5 years and 11 months), whereas the latter is available for children up to 83 months of age (6 years and 11 months). In our analysis, we report the results of a composite index obtained from these three subscales using principal components analysis.

2.3.4 Maternal mental health and endowments

We collected data on symptoms of depression using the Center for Epidemiologic Studies Depression Scale (CESD) (Knight et al. 1997), and collected measures of maternal distress with the Parenting Stress Index (PSI) (Abidin 1990). We also apply two scales of the Wechsler Adult Intelligence Scale (WAIS-IV) to caregivers (vocabulary and digit span). This allows us to control for maternal IQ, which is an important predictor of child cognitive skills. In addition, we also measure the caregiver's personality traits using the Big Five test (Goldberg 1993), which assesses extraversion, agreeableness, conscientiousness, openness and neuroticism.

⁹ A subset of children older than 36 months and younger than 71 months of age were administered as a result both the PLS-IV and the TEVI-R. The two measures align well for this subset.

Finally, we collected socio-economic data for all the household members including education attainment, age, labor and non-labor incomes, family composition, employment status, household wealth, access to health and community services, and health shocks.

3 Data

3.1 Baseline descriptive and sample balance

Table 1 describes key characteristics for our sample of 2,916 principal caregivers and their households. There are three sets of columns, one for each treatment arm. We show the mean of each variable and the number of observations in each group. The last two columns of the table display p-values of tests of whether the values of each variable are equal for the control group and for each of the treatment groups.

Caregivers are mostly mothers (94.8%), followed by grandmothers (3.6%). The father is the main caregiver for the child only in 1.2% of all households. This is consistent with what we see in the administrative records from the program. The average age of caregivers is 29 years old, with most caregivers being between 21 and 30 years of age. Moreover, 37.5% of caregivers are high school dropouts, and 16.3% have some level of tertiary education.

The intervention targets the most disadvantaged section of the population in Chile. Of the households in the sample, 52.1% belong to the bottom quintile and 80.2% belong to the bottom two quintiles of the household income distribution in the country (the definition of poverty in the table). Among the participants, 41.3% are single mother households, while the remaining are biparental non-extended families (consisting of father, mother, and children, but no other adults at home).

We find no significant differences between families across treatment arms at baseline whenever we test for equality of these characteristics individually or jointly (last line).

Table 1: Baseline balance, Caregiver and Household Characteristics

Variable	(1)		(2)		(3)		t-test	t-test
	N	Mean	N	Mean	N	Mean	p-value (1)-(2)	p-value (1)-(3)
Who is the caregiver (%)								
Mother	972	94.5%	972	95.2%	972	94.8%	0.538	0.840
Grandmother	972	4.0%	972	3.2%	972	3.5%	0.330	0.551
Other	972	0.5%	972	0.4%	972	0.4%	0.738	0.738
Father	972	0.9%	972	1.2%	972	1.3%	0.511	0.391
Caregiver's education (%)								
Primary	958	19.9%	958	22.5%	954	20.2%	0.163	0.873
Secondary incomplete	958	17.2%	958	17.1%	954	15.5%	0.952	0.313
Secondary complete	958	46.8%	958	44.3%	954	47.5%	0.271	0.753
Tertiary	958	16.1%	958	16.1%	954	16.8%	1.000	0.681
Single Mother	972	40.6%	972	41.2%	972	42.0%	0.818	0.550
Caregiver's age (%)								
15-20 years old	969	12.9%	969	15.1%	965	13.8%	0.169	0.568
21-30 years old	969	48.5%	969	48.6%	965	48.8%	0.964	0.893
31-40 years old	969	30.0%	969	26.8%	965	27.9%	0.119	0.296
41-50 years old	969	6.8%	969	7.5%	965	7.9%	0.538	0.370
>51 years old	969	1.8%	969	2.0%	965	1.7%	0.737	0.870
Hh'ld p.c. income (<40%)	972	80.0%	972	80.9%	972	79.7%	0.647	0.865
Hh'ld p.c. income (\$US)	972	209.7	972	209.7	972	209.7	0.354	0.997
F-test of joint significance (p-value)							0.538	0.840

Note: T-tests report comparisons between the control arm and NEP-B and NEP-I. Household per capita income reported in 2011 US dollars per month. Significance levels: * $p \leq 10\%$, ** $p \leq 5\%$. F-test for the joint significance across all variables is reported at the bottom.

In Appendix 1 (tables A1 to A6), we show that the sample is also balanced across treatment arms in terms of child characteristics, child development, parental beliefs, and parenting behaviors. We also show that child cognitive outcomes, such as language and executive function, are strongly positively correlated with maternal education, while children of more educated caregivers are less likely to exhibit behavioral problems (figures A1 and A2) (Fernald et al. 2012; Schady et al. 2014). These trends are mirrored by measures of parental behaviors and beliefs (figures A3 and A4). Relative to less educated caregivers, those who are more educated provide more cognitive stimulation to their children, are more nurturing, use less harsh disciplinary practices, have a higher perception of self-efficacy and see child rearing in a more authoritative way.

4 Empirical Strategy

We begin by estimating the impacts of offering NEP-B and NEP-I on parental behaviors and child development, what is usually referred to as the intention-to-treat parameter (ITT). Impacts on child outcomes are estimated using data at the child level, while impacts on parental outcomes such as behaviors, beliefs and well-being are estimated using data at the household level.

We use the following specification:

$$Y_{ic} = \alpha + \beta_1^{ITT} NEP_{ic}^B + \beta_2^{ITT} NEP_{ic}^I + \Gamma_c + X_{ic}\delta + \varepsilon_{ic} \quad (1)$$

where Y is the outcome of interest at endline, which varies by child / household i , in health center, c . NEP^B is an indicator which takes value 1 if the caregiver was invited to participate in NEP-B, and NEP^I is an indicator which takes value 1 if the caregiver was invited to NEP-I. X is a set of control variables including children's attributes such as sex and age, household characteristics such as family structure, household's per capita income, caregiver's education, and the outcome of interest at baseline. Γ_c are health center fixed effects which capture unobservable differences in program quality. Coefficients β_1 and β_2 are the ITT parameters of interest.

Since participation in NEP was voluntary, we complement the ITT analysis with the estimation of instrumental variables (IV) estimates of the impacts of NEP. If these impacts are heterogeneous in the population, these estimates are usually interpreted as Local Average Treatment Effects (LATE). The first and second stage equations for this estimator are:

$$T_{ic}^B = \alpha_0 + \alpha_1 NEP_{ic}^B + v_{ic}^B \quad (2)$$

$$T_{ic}^I = \gamma_0 + \gamma_1 NEP_{ic}^I + v_{ic}^I \quad (3)$$

$$Y_{1ic} = \beta_0 + \beta_1^{LATE} T_{ic}^B + \beta_2^{LATE} T_{ic}^I + \Gamma_c + X_{ic}\delta + \varepsilon_{ic} \quad (4)$$

where random assignment dummies NEP^B and NEP^I are used as instrumental variables for participation in each treatment arm, T^B and T^I (Basico and Intensivo, respectively).

5 Results

This section presents the estimated impacts of NEP-B and NEP-I on child outcomes and parenting attributes. Because the endline survey was conducted between 30 to 36 months after the end of the interventions for households in our sample, these can be interpreted as medium-term effects of the program. In this section we focus on the simplest specification where we only control for health center fixed effects. In tables A1-A4 in Appendix 2 we present specifications that include age, gender, household characteristics, maternal cognition and personality traits, and baseline outcomes as controls. Our results are robust to the inclusion of controls. We supplement standard inference procedures with multiple hypothesis testing (Romano and Wolf 2005), where all outcomes, including child development, parental behaviors and parental beliefs are considered simultaneously. Appendix 4 discusses in detail the construction of the final measures of child development, parenting, and beliefs used to estimate impacts in this section.

5.1 Intention-to-Treat

Table 2 shows our main ITT estimates of the impacts of NEP-B and NEP-I on standardized scores of child developmental outcomes. The results suggest that the offer of NEP-B improves receptive language (TEVI-R) by 0.076 SD (statistically significant at the 10% level) and NEP-I improves these by 0.10 SD (statistically significant at the 5% level). Accounting for multiple hypothesis testing, only the NEP-I impact remains statistically significant. In column 4 we show that we cannot reject the hypothesis of equal impacts of NEP-B and NEP-I.

Both NEP-B and NEP-I lead to an increase in the composite index of socio-emotional development of 0.064 SD and 0.132 SD, respectively. However, only NEP-I is statistically significant at the 5% level and robust to multiple hypothesis testing. These impacts are driven by changes in the subscales of interaction with adults and the social role, results reported in Table A6 in Appendix 2.

The interventions also lead to a decrease in behavioral problems (internalization and externalization) but the impacts are small and not statistically significant. The estimated impacts on executive function and sustained attention are also small and insignificant. These results are robust to the inclusion of additional controls, such as child age and sex, baseline outcomes, or caregiver socio-economic characteristics. This is shown in table A6 in Appendix 2.

Table 2: ITT estimates of child development outcomes

	Obs.	NEP-B	NEP-I	P value Test B=I
Receptive Language	2895	0.076* (0.044)	0.100**† (0.045)	0.623
Personal-Social Development: Composite Index	1532	0.064 (0.061)	0.132**† (0.062)	0.258
Behavioral problems: Externalization	1971	-0.022 (0.050)	-0.014 (0.050)	0.874
Behavioral problems: Internalization	1887	-0.028 (0.049)	-0.019 (0.049)	0.856
Executive Function	2879	-0.008 (0.044)	0.035 (0.045)	0.329
Sustained attention	2893	-0.035 (0.044)	0.009 (0.044)	0.318

Note: Each line reports estimates from a separate regression. Dependent variables are standardized to be mean 0, SD=1. Executive function is measured with the DCCS test, Receptive language is assessed using the TEVI-R test, personal social development is administered using the BDIST II, and behavioral problems (externalizing and internalizing behaviors) are assessed using the CBCL instrument. All regressions control for health center's fixed effects. *p<=10%, **p<=5% testing individual hypotheses, †p<=10% testing multiple hypotheses.

Table 3 presents ITT estimates for caregiver outcomes such as child-rearing practices, caregiver beliefs, attitudes and perceptions, and caregiver psychological well-being. Regarding parental behaviors, receiving an invitation to participate in NEP-I significantly improves the quality of the

home environment by 0.155 SD (measured by a combination of the FCI and HOME indicators, constructed using a principal component analysis). This estimate remains statistically significant even after accounting for multiple hypothesis testing. All results are robust to the inclusion of additional controls (see Table A8 in Appendix 2).

Statistically significant changes are found in Affection but not for Interaction, measured by the PBC nurturing scales. We find that NEP-I leads to a decrease in negative discipline (-0.077 SD), measured by the PBC discipline scale, but this estimate cannot be distinguished from zero with multiple hypothesis testing. Impacts of NEP-B on these variables are positive, but smaller in magnitude than those of NEP-I, and never statistically significant.

Regarding caregiver perceptions, we find that NEP-I significantly increases perceived self-efficacy (measured by the Parental Sense of Competence Scale) by 0.1 SD, and parental perceived impact of own behavior on child development by 0.1 SD, results which are robust to multiple hypothesis testing. We also find that perceived social support from friends and the community is also increased by 0.082 SD. While we find no statistically significant impacts of NEP-B or NEP-I on parental attitudes towards child-rearing measured through parental styles, and we see statistically significant impacts of NEP-I on the average age at which parents believe their children would be able to achieve key language developmental milestones in 0.1 SD.

Finally, we find no significant impacts of NEP in caregiver mental health (CESD) or stress (PSI).

5.2 Instrumental Variables

We now discuss IV estimates of the impact of participating in NEP-B and NEP-I on child and household outcomes. Based on information from administrative program records, the overall participation rates in NEP-B was 24.9% and in NEP -I was 30.8% among eligible individual. There is also imperfect compliance in the control group. The original plan was to start offering the program to the control group one year after the start of the study and given the delay of the baseline, part of the control group eventually received treatment. 4.8% of caregivers assigned to the control group were able to access the program, and attended at least one session of NEP-B. 5.0% of caregivers in the control group attended at least one session of NEP-I.

Table 3: ITT estimated parameters parental practices and parental beliefs

Dep. Var.:	Obs.	NEP-B	NEP-I	P-value test B=I
Parental Practices				
Home Index	2545	0.084 (0.072)	0.155**† (0.072)	0.313
PBC Affection	2545	0.042 (0.046)	0.085* (0.046)	0.348
PBC Interaction	2545	0.015 (0.046)	0.013 (0.046)	0.959
PBC Negative discipline	2545	-0.047 (0.047)	-0.077* (0.047)	0.516
PBC Positive discipline	2545	0.054 (0.048)	0.054 (0.048)	0.988
Parental Beliefs, Attitudes, Perceptions				
Perceived Self-efficacy	2543	0.037 (0.047)	0.100**† (0.047)	0.174
Perceived Parental Impact of own behavior on child development	2545	0.067 (0.046)	0.103**† (0.046)	0.429
Perceived Social Support - Family	2545	-0.078 (0.048)	0.005 (0.048)	0.076
Perceived Social Support - Friends	2545	0.071 (0.046)	0.082* (0.046)	0.797
Perceived Social Support – Others	2545	-0.014 (0.047)	0.013 (0.047)	0.554
Democratic style	2545	0.040 (0.048)	0.046 (0.048)	0.907
Authoritarian style	2545	0.026 (0.048)	-0.026 (0.048)	0.267
Permissive style	2545	-0.064 (0.048)	-0.018 (0.048)	0.323
Elicited Age High Investment Home Scenario	1487	-0.049 (0.050)	-0.064 (0.048)	0.763
Elicited Age Low Investment Home Scenario	1486	-0.057 (0.051)	-0.103* (0.054)	0.416
Psychological Well Being				
Parental Stress	2545	0.044 (0.047)	-0.011 (0.047)	0.231
Depression	2545	0.037 (0.047)	0.033 (0.047)	0.939

Note: Each line reports estimates from a separate regression. Dependent variables are standardized to be mean 0, SD=1. All regressions control for health center's fixed effects. Significance levels: *p<=10%, **p<=5% testing individual hypotheses, †p<=10% testing multiple hypotheses

5.2.1 Program Participation

Table 4 (columns 1 and 3) describes estimates of the first stage regressions when no covariates are included. Column 1 corresponds to equation (2), and Column 3 corresponds to equation (3). The impacts of being offered a slot in NEP-B and NEP-I on participation in these programs are 20.1% and 26.0%, respectively. Columns 2 and 4 of Table 4 add controls to the estimation of equations (2) and (3) and these estimates are hardly affected.¹⁰

Table 4 – Program take-up

	Participation NEP-B		Participation NEP-I	
	(1)	(2)	(3)	(4)
	Coef	Coef	Coef	Coef
	(SE)	(SE)	(SE)	(SE)
NEP-B	0.201***	0.200***		
	(0.012)	(0.012)		
NEP-I			0.260***	0.262***
			(0.013)	(0.013)
Controls	N	Y	N	Y
N	2,545	2,530	2,545	2,530

Note: Columns 1 and 3 control only for health center fixed effects. Columns 2 and 4 add households' socio-demographic characteristics as well as caregiver's labor status at baseline. Significance levels: * $p \leq 10\%$, ** $p \leq 5\%$, *** $p \leq 1\%$

When there is partial compliance only in the treatment group the IV estimate corresponds to the TT parameter. This is however not the case in our setting. As we mentioned above, there is some partial compliance in the control group as well. This is why we interpret our IV as a local average treatment effect (LATE), or the impact of the program on compliers.

¹⁰ In Appendix 2 table A10 we show the coefficients on the control variables. With regard to child characteristics, caregivers with a child between 25 and 36 months at baseline are 4.2% more likely to attend sessions in NEP Intensivo. We do not observe a significant association between household income and participation in NEP-B, but households belonging to the second income quantile are 3.9% more likely to attend NEP-I than those at the bottom of the income distribution. Interestingly, the likelihood of participation is higher among more educated caregivers in NEP-B, but it is not relevant to explain participation in NEP-I. Single mothers are 3.4% less likely to attend NEP-B and 3.2% less likely to attend NEP-I. Finally, caregivers that were employed at baseline are 3.0% less likely to participate in NEP-B and 2.5% less likely to participate in NEP-I. Taking the last two indicators together, the data suggest that there were important time constraints for participation among working caregivers with less support.

Finally, the average number of sessions attended by compliers was 5.68 sessions in NEP-B and 7.89 sessions in NEP-I. Therefore, the estimated impact on participants reported in tables 5 and 6 can be interpreted as the average impact of these number of sessions in each treatment arm.

5.2.2 Impact on program participants

Table 5 shows the IV estimates corresponding to equation (4). Participation in NEP-I improves receptive language performance by 0.432 SD and socio-emotional development by 0.540 SD, and NEP-B improves receptive language by 0.418 SD and socio-emotional development by 0.315 SD. However, only the former estimates are robust to multiple hypotheses testing. There are no statistically significant impacts on the remaining variables in the table, and the point estimates are also smaller in magnitude than those for language and socio-emotional. These results consistently mirror the ITT analysis that shows no significant changes in other child outcomes.¹¹

As we discussed above, our estimates indicate that the program was remarkably effective in improving language and socio-emotional outcomes of children in the medium term, given its low cost and low intensity. The estimates are striking both because of their size, and because the endline was collected so long after participants stopped participating in the program. As in the case of ITT estimates, when we add more controls in the 2SLS estimation, the coefficients for child outcomes remain fairly stable, and standard errors hardly change either (Table A7 in Appendix 2).

¹¹ These impacts are subject to the accuracy of the administrative data of participation rates. When a second source of participation based on survey records (from the endline) is considered, participation is significantly higher: 44% in NEP-B and 50% in NEP-I. Administrative records are usually judged to be superior to survey answers, but they might still have some degree of under-reporting if for example the facilitators at the health center fail to register some sessions electronically. Household surveys might overreport participation because they are subject to recall and desirability. Using self-reported participation records we get lower IV estimates of 0.2 SD in the intensive arm.

Table 5: Effect of participation in NEP on child development outcomes

	Obs.	NEP-B	NEP-I	P value Test B=I
Receptive Language	2895	0.418* (0.242)	0.432**† (0.189)	0.963
Personal-Social Development: Composite Index	1509	0.315 (0.304)	0.540**† (0.261)	0.361
Behavioral problems: Externalization	1971	-0.118 (0.271)	-0.068 (0.212)	0.814
Behavioral problems: Internalization	1887	-0.150 (0.268)	-0.088 (0.203)	0.770
Executive Function	2879	-0.038 (0.237)	0.130 (0.186)	0.374
Sustained Attention	2893	0.009 (0.044)	0.011 (0.185)	0.295

Note: Each line reports estimates from a separate regression. All regressions control for health center’s fixed effects. Significance levels: * $p \leq 10\%$, ** $p \leq 5\%$ testing individual hypotheses, † $p \leq 10\%$ testing multiple hypotheses.

Our estimates of the impacts of NEP on parental behaviors and home environments suggest that, in spite of their large magnitude, the medium-term impacts on language outcomes are plausible, because they are mirrored by sustained changes in parenting behaviors and home environments for participating households. Table 6 shows that NEP-I significantly improves the HOME index in the medium term by 0.66 SD (also robust to multiple hypothesis testing), and decreases negative discipline by 0.33 SD. NEP-I also significantly improves the Affection index measured by the PBC nurturing scale, by 0.36 SD. These are substantial impacts.

Regarding parental beliefs, we observe a positive and large impact of NEP-I on parental self-efficacy of 0.411 SD and on parental perceived impact of own behavior on child development by 0.446 SD (both robust to multiple hypothesis testing). We also observe an improvement of perceived social support of 0.37 SD. There are no significant changes in parental styles.

Finally, parents receiving the NEP-I treatment expect their children to reach particular developmental outcomes at a younger age than those in the control group about the age at which children. This is especially true in a “low-investment” home scenario (by 0.39 SD).

Again, our IV results in parental outcomes are robust to the inclusion of additional controls such household socio-demographic characteristics, caregiver endowments and outcomes at baseline, as shown in Table A9 in Appendix 2.

The estimates of the impacts of NEP-I on parental behaviors and home environments (in particular, on the HOME index) are as remarkable as the estimates on child language reported above, precisely for the same reasons already discussed: they are large in magnitude (especially when we take into account the low costs of the program), and they are long lasting. The fact that we observe

these impacts occurring simultaneously for more than one measure strengthens our confidence that this program leads to improved home environments.

We cannot fully rule out that the medium-term impacts of NEP on home environments are due exclusively to what parents learned in the 6 to 8 NEP sessions in which they participated. For example, participation in NEP may encourage parents to look for additional parenting programs in the future, in which case what we observe is the compounded impact of multiple parenting programs. It is also possible that NEP helps establish a strong network of neighbors with similar parenting concerns, leading up to the establishment of informal support groups in the community. In this case, our estimates of NEP would capture the impacts of these networks as well. In sum, the overall impacts of NEP may go well beyond whatever parenting training parents benefit from in the 6-8 sessions they participate in the program.

Finally, we also examined heterogeneous impacts of NEP on our main outcomes of interest along two dimensions: caregiver SES and child gender. Overall, we did not find strong evidence of heterogeneity across these dimensions, so we report all these results in Appendix 5. If anything, our results suggest that there are stronger impacts girls, and for children from more disadvantaged families.

Table 6: Effect of participation in NEP on parental practices and beliefs

	Obs.	NEP-B	NEP-I	P value for B=I
Parental Practices				
Home Index	2545	0.464 (0.396)	0.659**† (0.311)	0.532
PBC Affection	2545	0.234 (0.254)	0.358* (0.199)	0.536
PBC Interaction	2545	0.082 (0.251)	0.059 (0.197)	0.907
PBC Negative discipline	2545	-0.262 (0.257)	-0.332* (0.201)	0.731
PBC Positive discipline	2545	0.294 (0.262)	0.246 (0.205)	0.817
Parental Beliefs, Attitudes, Perceptions				
Perceived Self-efficacy	2543	0.205 (0.256)	0.411**† (0.202)	0.311
Perceived Parental Impact of own behavior on child development	2545	0.372 (0.254)	0.446**† (0.199)	0.714
Perceived Social Support – Family	2545	-0.423 (0.263)	-0.033 (0.206)	0.059
Perceived Social Support – Friends	2545	0.389 (0.255)	0.367* (0.200)	0.912
Perceived Social Support - Others	2545	-0.075 (0.257)	0.042 (0.201)	0.560
Democratic style	2545	0.222 (0.264)	0.205 (0.207)	0.936
Authoritarian style	2545	0.140 (0.262)	-0.084 (0.205)	0.277
Permissive style	2545	-0.350 (0.262)	-0.112 (0.206)	0.249
Elicited Age High Investment Home Scenario	1487	-0.262 (0.256)	-0.248 (0.179)	0.947
Elicited Age Low Investment Home Scenario	1486	-0.305 (0.295)	-0.387* (0.205)	0.736
Psychological Well Being				
Parental Stress	2545	0.238 (0.255)	-0.013 (0.200)	0.211
Depression	2545	0.201 (0.255)	0.153 (0.200)	0.812

Note: Each line reports estimates from a separate regression. All regressions control for health center's fixed effects. Significance levels: *p<=10%, **p<=5% testing individual hypotheses, †p<=10% testing multiple hypotheses.

There is roughly 19% attrition from the baseline to the endline surveys, which is quite substantial. We discuss attrition in detail in Appendix 3. We show that there is slightly more attrition in the control than either of the treatment groups, but that the differences are small. In addition, we present two estimators of our main equations (with special attention to child outcomes) which account for selective attrition. One is a control function approach where interviewer fixed effects are used as an instrument for attrition. The second censors the outcomes of interest (say a test score) at different values and examines how estimates change, as in Angrist, Bettinger, and Kremer (2006). Our results are robust to these corrections for attrition.

5.3 Mediation Analysis

Our results above show that as a consequence of participation in NEP there are sustained improvements in both child vocabulary and socio-emotional scores and home environments. Given that the duration of the program is quite short, and child outcomes are measured 2 to 3 years after the end of the program, it is reasonable to think that any program impacts operate primarily by changing in parenting behaviors in the long run. Given the richness of our data on parenting behaviors and home environments, in this section we investigate if this idea is consistent with our data.

In this section we present a standard mediation analysis to examine to what extent the impacts on NEP on the vocabulary and socio-personal development scores of children can be explained by the impacts of NEP on parental behaviors and beliefs. The assumptions under which one can decompose treatment effects estimates into different components are however very strong. This means that the results in this section can only be interpreted as suggestive evidence of the importance of these mediators.

In a standard mediation model where the outcome of interest is Y and the mediating factor (observed measured input) is M (it can be a vector of factors), the goal is to separately identify the intervention's total indirect effect ($a * b$) from the direct effect (c) from the following model:

$$M = a_0 + aD + u \quad (5)$$

$$Y = b_0 + bM + cD + \varepsilon \quad (6)$$

where a is the ITT estimate of NEP on a particular mediator (practices and beliefs), and b the marginal effect of mediator M on the outcome. The outcomes we consider is the child's vocabulary score and the socio-personal development index, and the vector of potential mediators includes quality of the home environment, nurturing and disciplinary strategies, self-efficacy, perceived social support, and parental beliefs about the process of child development. In the case of NEP-B, the program was unable to shift significantly any intermediate indicators, so we focus on NEP-I.

We estimate the model in steps using a Monte Carlo simulation approach following Campos et al. (2017). First, we estimate the coefficient a by regressing the effect of treatment assignment on each mediator (eq. 5). Second, we obtain estimates of b from a regression of child language on treatment status (as in the ITT equation, controlling for child and household characteristics and health center fixed effects) and add one particular mediator at a time. We then compute the 95%

Monte Carlo confidence intervals for $(a * b)$ the indirect effect based on a very large number of repetitions. A confidence interval that does not include zero indicates a significant indirect effect of that particular mediating variable on child outcomes. Finally, in order to assess the Monte Carlo confidence intervals for the total indirect effect we include all the relevant mediators (the significant a paths) and the mediators that resulted to be significant individually (the b paths) in the same regression model.

Table 7 describes our main results for language development. Column 1 reports the ITT coefficients of the impact of the program on language (from table 6). Columns (2)-(4) add one significant mediator at the time to the model; column (5) adds all intermediate outcomes that are significantly shifted by the NEP-I. The mediators (both in terms of behaviors and in terms of beliefs, attitudes) are jointly significant in explaining the main outcome of interest.

The home environment, the index of nurturing practices with children, the index of negative discipline, perceived self-efficacy, perceived social support by friends and perceived parental impact in child development are significantly affected by the intervention. The direct impact estimate is 0.100, which declines to 0.087 when we add the significant mediators (column 5) and becomes only marginally significant. This means that we can at most explain about 13% of the effect of NEP-I on child language through the impact of NEP-I in these potential mediators.

Table 8 examines the mediating factors of impacts found in socio-emotional development. The overall impact of NEP-I is 0.132 SD, which is significant at the 5% level, but after adding all significant mediators it declines to 0.084 SD and is no longer statistically significant. That is, mediating factors explain up to 36% of the treatment effects, of which the quality of the home environment explains half of it, and the other half is explained by nurturing and disciplinary practices.

In sum, we can only partially account for the main channels through which NEP is affecting child vocabulary in the medium run, but our rich data set is able to explain much better the mediating pathways of our impacts found in socio-emotional development. In the case of language, the impacts of NEP come through other unobserved channels, not discussed in this analysis. It is possible that NEP led to changes in other dimensions of the home environments, uncorrelated with the ones we observe. Or it is possible that NEP encouraged parents to search for subsequent early childhood programs publicly or privately provided. However, whatever the explanation is, it is unobservable, and unfortunately we can only speculate about it.

Table 7: Mediation Analysis Receptive Language

	(1)	(2)	(3)	(4)	(5)
	Main ITT	+HOME	+PBC nurturing	+PBC discipline	+significant mediators
NEP Intensive	0.076* (0.045)	0.071 (0.044)	0.068 (0.044)	0.069 (0.044)	0.075* (0.045)
Home Index+	0.100** (0.045)	0.090** (0.045)	0.086* (0.045)	0.088* (0.045)	0.087* (0.045)
PBC nurturing		0.061*** (0.013)	0.052*** (0.013)	0.053*** (0.013)	0.052*** (0.014)
PBC discipline			0.042** (0.020)	0.041** (0.020)	0.042** (0.021)
Perceived self-efficacy				0.024 (0.019)	0.032 (0.022)
Perceived social support - friends					0.041* (0.023)
Perceived parental impact					-0.023 (0.020)
Confidence Intervals for the joint significant indirect effect ($a * b$)					
Lower Bound		0.001	0.002	-0.003	-0.020
Upper Bound		0.022	0.022	0.023	0.020
No. Observations	1908	1908	1908	1908	1885
p-value joint significance behaviors		0.000	0.000	0.000	0.000
p-value joint significance beliefs					0.0794

Note: Each column reports estimates from a separate regression. Estimates control for health center's fixed effects. Significance levels: *p<=10%, **p<=5%, ***p<=1%.

Table 8: Mediation Analysis Socio-emotional development

	(1)	(2)	(3)	(4)	(5)
	Main ITT	+HOME	+PBC nurturing	+PBC discipline	+significant mediators
NEP Intensive	0.132** (0.062)	0.106* (0.060)	0.101* (0.059)	0.085 (0.058)	0.084 (0.058)
Home Index+		0.136*** (0.017)	0.091*** (0.017)	0.089*** (0.017)	0.079*** (0.017)
PBC nurturing			0.232*** (0.031)	0.241*** (0.031)	0.229*** (0.031)
PBC discipline				-0.158*** (0.025)	-0.098*** (0.030)
Perceived self-efficacy					0.097*** (0.031)
Perceived social support - friends					0.078*** (0.028)
Perceived parental impact					-0.002 (0.027)
Confidence Intervals for the joint significant indirect effect ($a * b$)					
Lower Bound		0.002	0.009	0.017	0.016
Upper Bound		0.042	0.006	0.078	0.078
No. Observations	1908	1908	1908	1908	1885
p-value joint significance behaviors		0.000	0.000	0.000	0.000
p-value joint significance beliefs					0.000

Note: Each column reports estimates from a separate regression. Estimates control for health center's fixed effects. Significance levels: * $p \leq 10\%$, ** $p \leq 5\%$, *** $p \leq 1\%$. Column (6) also controls for a socio-emotional score, PBC interaction, PBC Discipline, Perceived social support from family, friends and others, Parenting Styles (democratic and authoritarian), Parenting Stress, Depression.

6 Conclusion

There is a large consensus across disciplines on the importance of high-quality interventions during the early years, a period in which critical cognitive and socio-emotional development processes are consolidated, with long-term implications for adulthood. Human capital investments during early childhood are not only important on the grounds of efficiency, given that earlier interventions have larger returns in the long-term, but also from the point of view of equity, as early childhood interventions are likely to reduce socio-economic gaps and the intergenerational transmission of poverty. Parents and caregivers play a key role in home stimulation during the early years, which is fundamental for healthy child development and is crucial to close early socio-demographics gaps in skills development.

This paper studies the medium-term results of a large-scale parenting program in Chile. The intervention, known as *Nadie es Perfecto* or NEP, provides information and support to parents and caregivers of the poorest and more disadvantaged groups, using a semi-structured curriculum where trained and certified facilitators who encourage group discussions about parental needs and concerns. The method behind the policy is based on experiential learning. Parents and caregivers share and learn from other parents' experiences and discuss the challenges of parenting that prevent the adoption of new strategies at home. The main objective of NEP is to change parental beliefs about their role in nurturing and facilitating the adoption of positive practices that reflect an improved parent-child interaction which, in turn, translates into child developmental outcomes.

Our results show sustained effects on parenting beliefs, practices and child outcomes three years after the intervention ended. The impact of the offer of the program show a significant positive effect of 0.1 SD in receptive language and an impact of 0.13 SD in socio-emotional development. The child outcome effects are mirrored by sustained changes in parenting practices: an increase of 0.16 SD in cognitive stimulation, a decrease of 0.08 SD in the use of negative disciplinary practices and a suggestive positive effect on affection. The intervention is able to shift parental beliefs and expectations. We observe a significant improvement of 0.1 SD in parental perceived self-efficacy, of 0.1 SD in the perceptions that parent have on how their behavior impacts their children's socio-emotional development and an increase of 0.08 SD in perceived social support. The effective attendance to group sessions ranged between 25% in NEP Basico to 31% in NEP Intensivo. When accounting for take-up, the outcomes of the intervention among participants suggest a substantial improvement of 0.43 SD in language development, 0.54 SD in the social development and similar effects on nurturing practices and parental beliefs.

Our results suggest that NEP seems to operate by changing parental beliefs and expectations and by improving positive parenting strategies with children. Results from a mediation analysis suggest that these factors do a good job in explaining NEP impacts on socio-emotional outcomes, but play only a limited role in mediating NEP impacts on receptive language. Among all the potential mediators that we found the intervention was impactful, the quality of the home environment, parental nurturing and disciplinary practices, the caregiver's perceived self-efficacy, the caregiver's perceived social support and the caregiver's perceived role in influencing child development were found to have a statistically significant mediating role of the NEP in child language and socio-emotional development, jointly explaining about 13% and 36% of the total effect of the intervention in these outcomes, respectively.

References

- Abidin, Richard R. 1990. *Parenting Stress Index:(Short Form)*. Pediatric Psychology Press Charlottesville, VA.
- Aboud, Frances E, and Sadika Akhter. 2011. "A Cluster-Randomized Evaluation of a Responsive Stimulation and Feeding Intervention in Bangladesh." *Pediatrics* 127(5): e1191 e1197.
- Aboud, Frances E, and Aisha K Yousafzai. 2015. "Global Health and Development in Early Childhood." *Annual Review of Psychology* 66: 433–57.
- Achenbach, Thomas M, and Thomas M Ruffle. 2000. "The Child Behavior Checklist and Related Forms for Assessing Behavioral/Emotional Problems and Competencies." *Pediatrics in Review* 21(8): 265–71.
- Angrist, Joshua, Eric Bettinger, and Michael Kremer. 2006. "Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia." *American Economic Review* 96(3): 847–62.
- Attanasio, Orazio et al. 2018. "Early Stimulation and Nutrition: The Impacts of a Scalable Intervention." NBER Working Paper (25059).
- Attanasio, Orazio P et al. 2014. "Using the Infrastructure of a Conditional Cash Transfer Program to Deliver a Scalable Integrated Early Child Development Program in Colombia: Cluster Randomized Controlled Trial." *BMJ* 349: g5785.
- Bandura, Albert. 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Prentice-Hall, Inc.
- . 1995. *Self-Efficacy in Changing Societies*. Cambridge University Press.
- Baumrind, Diana. 1968. "Authoritarian vs. Authoritative Parental Control." *Adolescence*.
- Black, Maureen M et al. 2017. "Early Childhood Development Coming of Age: Science through the Life Course." *The Lancet* 389(10064): 77–90.
- Blair, Clancy, and Rachel Peters Razza. 2007. "Relating Effortful Control, Executive Function, and False Belief Understanding to Emerging Math and Literacy Ability in Kindergarten." *Child Development* 78(2): 647–63.
- Bock, R D, M F Zimowski, W J Van Der Linden, and R K Hambleton. 1996. "Multiple Group IRT." *Handbook of Modern Item Response Theory*.
- Boivin, Michel et al. 2005. "The Genetic-Environmental Etiology of Parents' Perceptions and Self-Assessed Behaviours toward Their 5-Month-Old Infants in a Large Twin and Singleton Sample." *Journal of Child Psychology and Psychiatry* 46(6): 612–30.
- Bradley, Robert H, and Bettye M Caldwell. 1984. "The Relation of Infants' Home Environments to Achievement Test Performance in First Grade: A Follow-up Study." *Child Development*: 803–9.
- Briscoe, Ciara, and Frances Aboud. 2012. "Behaviour Change Communication Targeting Four Health Behaviours in Developing Countries: A Review of Change Techniques." *Social Science & Medicine* 75(4): 612–21.
- Britto, Pia R et al. 2016. "Advancing Early Childhood Development: From Science to Scale 2 Nurturing Care: Promoting Early Childhood Development." *safety (eg, routines and protection from harm)* 3: 4.
- Campbell, Frances A, and Craig T Ramey. 1994. "Effects of Early Intervention on Intellectual and Academic Achievement: A Follow-up Study of Children from Low-Income Families." *Child Development* 65(2): 684–98.
- Campos, Francisco et al. 2017. "Teaching Personal Initiative Beats Traditional Training in Boosting Small Business in West Africa." *Science* 357(6357): 1287–90.
- Cu 2013. "Eliciting Maternal Expectations about the Technology of Cognitive Skill Formation". NBER Working Paper No. 19144.
- Cutrona, Carolyn E, and Beth R Troutman. 1986. "Social Support, Infant Temperament, and Parenting Self-Efficacy: A Mediational Model of Postpartum Depression." *Child Development*: 1507–18.

- Desjardins, Christopher D et al. 2018. "A Generalized Rasch Model for Manifest Predictors A Generalized Rasch Model for Manifest Predictors." In *Handbook of Educational Measurement and Psychometrics Using R*, Lincolnwood Hoboken, NJ, 1–29.
- Doyle, Orla et al. 2013. "Measuring Investment in Human Capital Formation: An Experimental Analysis of Early Life Outcomes." NBER Working Paper No. 19316.
- Echeverria, Max S, Maria Olivia Herrera, and J Teresa Segure. 2002. *TEVI-R: Test de Vocabulario En Imágenes*. Universidad de Concepción.
- Engle, Patrice L et al. 2007. "Strategies to Avoid the Loss of Developmental Potential in More than 200 Million Children in the Developing World." *The Lancet* 369(9557): 229–42.
- . 2011. "Strategies for Reducing Inequalities and Improving Developmental Outcomes for Young Children in Low-Income and Middle-Income Countries." *The Lancet* 378(9799): 1339–53.
- Fernald, Lia C H, Patricia Kariger, Melissa Hidrobo, and Paul 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): 17273–80.
- Fox, Robert. 1994. "Parent Behavior Checklist."
- Gertler, Paul et al. 2014. "Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica." *Science* 344(6187): 998–1001.
- Goldberg, Lewis R. 1993. "The Structure of Phenotypic Personality Traits." *American Psychologist* 48(1): 26.
- Gowani, Saima, Aisha K Yousafzai, Robert Armstrong, and Zulfiqar A Bhutta. 2014. "Cost Effectiveness of Responsive Stimulation and Nutrition Interventions on Early Child Development Outcomes in Pakistan." *Annals of the New York Academy of Sciences* 1308(1): 149–61.
- Grantham-McGregor, Sally M, Christine A Powell, Susan P Walker, and John H Himes. 1991. "Nutritional Supplementation, Psychosocial Stimulation, and Mental Development of Stunted Children: The Jamaican Study." *The Lancet* 338(8758): 1–5.
- Hamadani, Jena D et al. 2010. "Use of Family Care Indicators and Their Relationship with Child Development in Bangladesh." *Journal of Health, Population, and Nutrition* 28(1): 23.
- Hamadani, Jena D, Syed N Huda, Fahmida Khatun, and Sally M Grantham-McGregor. 2006. "Psychosocial Stimulation Improves the Development of Undernourished Children in Rural Bangladesh." *The Journal of Nutrition* 136(10): 2645–52.
- Heckman, James J. 2006. "Skill Formation and the Economics of Investing in Disadvantaged Children." *Science* 312(5782): 1900–1902.
- . 2017. "An Analysis of the Memphis Nurse-Family Partnership Program." NBER Working Paper No. 23610.
- Heckman, James J, and Dimitriy V Masterov. 2007. "The Productivity Argument for Investing in Young Children." *Applied Economic Perspectives and Policy* 29(3): 446–93.
- Heming, G, P A Cowan, and C P Cowan. 1990. "Ideas about Parenting." *Handbook of Family Measurement Techniques*: 362–63.
- Kagitcibasi, Cigdem et al. 2009. "Continuing Effects of Early Enrichment in Adult Life: The Turkish Early Enrichment Project 22 Years Later." *Journal of Applied Developmental Psychology* 30(6): 764–79.
- Knight, Robert G, Sheila Williams, Rob McGee, and Susan Olan. 1997. "Psychometric Properties of the Centre for Epidemiologic Studies Depression Scale (CES-D) in a Sample of Women in Middle Life." *Behaviour Research and Therapy* 35(4): 373–80.
- Kolb, David A. 2014. *Experiential Learning: Experience as the Source of Learning and Development*. FT press.
- Lu, Irene R R, D Roland Thomas, and Bruno D Zumbo. 2005. "Embedding IRT in Structural Equation Models: A Comparison with Regression Based on IRT Scores." *Structural Equation Modeling* 12(2): 263–77.
- Maccoby, Eleanor E, John A Martin, P H Mussen, and E Mavis Hetherington. 1983. "Handbook of Child

- Psychology: Vol. 4. Socialization, Personality, and Social Development.” PH Mussen (Series Ed.): 1–101.
- Nahar, Baitun et al. 2012. “Effects of a Community-Based Approach of Food and Psychosocial Stimulation on Growth and Development of Severely Malnourished Children in Bangladesh: A Randomised Trial.” *European Journal of Clinical Nutrition* 66(6): 701–9.
- Ohan, Jeneva L, Debbie W Leung, and Charlotte Johnston. 2000. “The Parenting Sense of Competence Scale: Evidence of a Stable Factor Structure and Validity.” *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement* 32(4): 251.
- Richter, Linda M. et al. 2017. “Investing in the Foundation of Sustainable Development: Pathways to Scale up for Early Childhood Development.” *The Lancet* 389(10064): 103–18.
- Ringwalt, Sharon. 2008. “Developmental Screening and Assessment Instruments with an Emphasis on Social and Emotional Development for Young Children Ages Birth through Five.” National Early Childhood Technical Assistance Center (NECTAC).
- Romano, Joseph P, and Michael Wolf. 2005. “Stepwise Multiple Testing as Formalized Data Snooping.” *Econometrica* 73(4): 1237–82.
- Schady, Norbert et al. 2015 “Wealth Gradients in Early Childhood Cognitive Development in Five Latin American Countries.” *The Journal of Human Resources*, 50(2): 446-463
- Singla, Daisy R, Elias Kumbakumba, and Frances E Aboud. 2015. “Effects of a Parenting Intervention to Address Maternal Psychological Wellbeing and Child Development and Growth in Rural Uganda: A Community-Based, Cluster-Randomised Trial.” *The Lancet Global Health* 3(8): e458-e469.
- Vazir, Shahnaz et al. 2013. “Cluster-Randomized Trial on Complementary and Responsive Feeding Education to Caregivers Found Improved Dietary Intake, Growth and Development among Rural Indian Toddlers.” *Maternal & Child Nutrition* 9(1): 99–117.
- Yousafzai, Aisha K et al. 2014. “Effect of Integrated Responsive Stimulation and Nutrition Interventions in the Lady Health Worker Programme in Pakistan on Child Development, Growth, and Health Outcomes: A Cluster-Randomised Factorial Effectiveness Trial.” *The Lancet* 384(9950): 1282–93.
- Zelazo, Philip David. 2006. “The Dimensional Change Card Sort (DCCS): A Method of Assessing Executive Function in Children.” *Nature Protocols - Eletronic Edition* 1(1): 297.

Appendix 1: Baseline Sample Characteristics (extended tables)

Table A1: Baseline balance, Child Characteristics

Variable	(1) Control		(2) NEP-B		(3) NEP-I		t-test p-value	t-test p-value
	N	Mean	N	Mean	N	Mean	(1)-(2)	(1)-(3)
Boys	1214	53.5%	1193	52.6%	1190	53.9%	0.658	0.841
Age in months								
0-12	1214	25.9%	1193	22.5%	1190	24.0%	0.058*	0.300
13-24	1214	22.0%	1193	24.4%	1190	23.4%	0.163	0.423
25-36	1214	16.0%	1193	19.6%	1190	17.8%	0.020**	0.230
37-48	1214	18.5%	1193	15.3%	1190	17.2%	0.042**	0.433
49-60	1214	10.5%	1193	11.7%	1190	11.6%	0.353	0.411
61-72	1214	7.1%	1193	6.2%	1190	5.8%	0.386	0.200
Birth Order								
First	1130	55.2%	1094	57.9%	1093	54.5%	0.209	0.743
Second	1130	29.9%	1094	28.7%	1093	32.2%	0.531	0.243
Third or more	1130	14.9%	1094	13.4%	1093	13.3%	0.334	0.278
F-test of joint significance (p-value)							0.061*	0.581

Note: T-tests report comparisons between the control arm against NEP-B and NEP-I. Significance levels: * $p \leq 10\%$, ** $p \leq 5\%$, *** $p \leq 1\%$. F-test for the joint significance across all variables is reported at the bottom.

Table A1 above presents the main descriptive statistics for the sample of 3,597 children participating in the evaluation at baseline, respectively. Among them, 53.4% are males, and the average age was 27.96 months. About half of the sample children in the study are below 2 years old at baseline (47.4%), more than half are the first child born, and 1/3 of them are the second born.

We do not find significant differences across gender or in the order of the child among all siblings. However, there are some small but statistically significant differences in children's age when we compare NEP Basico with the Control group. In order to correct for potential biases due to imbalance in age groups, we report treatment effects that also control for age and gender of the child.

Table A2 shows means and standard deviations of children's performance in both receptive and expressive language development measured at baseline using the PLSIV scale. There are no significant differences across treatment arms either when we use global T scores, or when we use the T scores for the receptive and expressive language sub-dimensions. Using the global T scores to diagnose developmental delays, we find that 16.7% of our sample between 3 months and 5 years old are diagnosed with some degree of delay, and that 5.8% of children are diagnosed with a clinical delay.

Table A2: Baseline Balance: Child Receptive and Expressive Language

	(1)	(2)	(3)	t-test	t-test
	Control	NEP-B	NEP-I	p-value	p-value
	Mean/SE	Mean/SE	Mean/SE	(1)-(2)	(1)-(3)
Language score (PLSIV)					
Global score	99.453 (0.529)	100.350 (0.539)	99.553 (0.544)	0.235	0.895
Receptive Language score	101.146 (0.544)	102.225 (0.545)	101.157 (0.552)	0.161	0.988
Expressive Language score	97.574 (0.491)	98.063 (0.516)	97.732 (0.524)	0.492	0.826
Diagnosis (Based on Global score)					
Clinical range (%)	0.058	0.047	0.070	0.268	0.266
Risk (%)	0.114	0.114	0.097	0.983	0.211
Normal (%)	0.828	0.839	0.833	0.518	0.763
Observations	1089	1060	1049		
F-test of joint significance (p-value)				0.719	0.388

Note: T-tests report comparisons between the control arm against NEP-B and NEP-I. Significance levels: *p<=10%, **p<=5%, ***p<=1%. F-test for the joint significance across all variables is reported at the bottom.

Table A3 describes child internalizing and externalizing behavioral problems reported by caregiver using the Child Behavior Checklist scale. The survey is applied to mothers of all children between 18 months and 5 years old. In our sample, 28.5% of children shows some mild or severe level of alteration (27.3% internalizing and 19.5% externalizing). There are no significant differences in scores across groups for any sub-dimension, and while there is a marginally significant difference between the Control group and NEP Basic in the percentage of children with moderate risk, the joint test across variables suggests a very low risk of sample imbalance.

Table A3: Baseline Balance: Child Maladaptive Behavior

Maladaptive Behavior, CBCL test	(1) Control	(2) NEP-B	(3) NEP-I	t-test p-value	t-test p-value
Variable	Mean/SE	Mean/SE	Mean/SE	(1)-(2)	(1)-(3)
T score, Global	56.828 (0.451)	57.094 (0.448)	56.316 (0.468)	0.676	0.430
T score, Internalization	56.130 (0.449)	56.290 (0.450)	55.939 (0.456)	0.802	0.765
T score, Externalization	54.990 (0.412)	55.299 (0.411)	54.595 (0.433)	0.595	0.508
Diagnosis (based on Global score)					
Clinical range (%)	0.155	0.135	0.143	0.270	0.518
Risk (%)	0.125	0.156	0.138	0.083*	0.466
Normal (%)	0.720	0.709	0.719	0.635	0.972
Observations	774	769	754		
F-test of joint significance (p-value)				0.516	0.681

Note: T-tests report comparisons between the control arm against NEP-B and NEP-I. Significance levels: * $p \leq 10\%$, ** $p \leq 5\%$, *** $p \leq 1\%$. F-test for the joint significance across all variables is reported at the bottom.

Table A4 shows the Dimensional Card Sort scale (DCCS) measure of executive functions performance in children older than 24 months old. In the test, if the child does not pass the first stage, she cannot be evaluated, which means that her performance is too low to be measured by the scale. If the child passes the first stage, she is evaluated as “Normal” if she completes the task, or “Altered” if she leaves the task incomplete. For example, the table shows that the proportion of children with “Altered” results out of those who passed the first stage is about 19.7% for children in the 36-47 months range, and 17.6% for children in the 48-59 months range and 11.4% for older children. We did not find any significant differences in the diagnostic across groups, except for the percentage of children with altered scores in the age groups 24-25 and 60-72 months, and the percentage of children that fail to pass the pre-change in the age group 60-72 months. Once again, the sample is fairly balanced across the three treatment arms.

Table A4: Baseline Balance: Executive function performance

Executive Function (DCCS)	(1) Control	(2) NEP-B	(3) NEP-I	t-test (1)-(2)	t-test (1)-(3)
Variable	Mean/SE	Mean/SE	Mean/SE	p-value	p-value
24-35 months					
Score	6.237 (0.265)	6.126 (0.255)	6.064 (0.267)	0.764	0.646
Fail to pass pre-stage (%)	0.711	0.722	0.706	0.800	0.920
Altered (%)	0.137	0.078	0.123	0.051	0.674
Normal (%)	0.153	0.200	0.172	0.208	0.612
Observations	190	230	204		
36-47 months					
Score	9.162 (0.205)	8.939 (0.232)	9.039 (0.218)	0.471	0.682
Fail to pass pre-stage (%)	0.332	0.356	0.305	0.619	0.560
Altered (%)	0.179	0.194	0.202	0.700	0.554
Normal (%)	0.489	0.450	0.493	0.439	0.937
Observations	222	180	203		
48-59 months					
Score	10.881 (0.180)	10.396 (0.217)	10.428 (0.213)	0.090	0.109
Fail to pass pre-stage (%)	0.071	0.072	0.116	0.987	0.219
Altered (%)	0.143	0.230	0.152	0.070	0.832
Normal (%)	0.786	0.698	0.732	0.104	0.310
Observations	126	139	138		
60-72 months					
Score	11.024 (0.230)	11.284 (0.223)	10.522 (0.341)	0.421	0.211
Fail to pass pre-stage (%)	0.035	0.054	0.116	0.568	0.054
Altered (%)	0.129	0.068	0.145	0.198	0.782
Normal (%)	0.835	0.878	0.739	0.444	0.145
Observations	85	74	69		

Note: T-tests report comparisons between the control arm against NEP-B and NEP-I. Significance levels: *p<=10%, **p<=5%, ***p<=1%.

Table A5 describes parental beliefs, psychosocial well-being, and investments in children. The scale Ideas About Parenting, measuring parenting styles, does not show significant differences across treatment arms. We also do not find significant differences in parental perceived self-efficacy, or in perceived social support. Finally, we do not find significant differences our

measures of parental investments in children using a measure of home environments based on the Family Care Indicators (FCI), or the sub-scales of Nurturing and Discipline from the Parenting Behavior Checklist.

Table A5: Baseline balance: parental beliefs, mental health and investments in children

Parental Indicators	(1)	(2)	(3)	t-test	t-test
	Control	NEP-B	NEP-I	p-value	p-value
Variable	Mean/SE	Mean/SE	Mean/SE	(1)-(2)	(1)-(3)
Authoritative style (IRT score)	-0.272 (0.021)	-0.287 (0.021)	-0.275 (0.021)	0.625	0.909
Authoritarian style (IRT score)	0.411 (0.024)	0.374 (0.024)	0.388 (0.025)	0.276	0.508
Permissive style (IRT score)	-0.538 (0.014)	-0.511 (0.015)	-0.539 (0.014)	0.189	0.963
Perceived Self-efficacy	64.220 (0.302)	64.173 (0.298)	64.545 (0.299)	0.911	0.444
Perceived Social Support	2.920 (0.054)	2.903 (0.055)	2.852 (0.054)	0.825	0.375
Parental Stress	29.943 (0.373)	30.361 (0.370)	29.676 (0.368)	0.427	0.610
Depression	40.222 (0.406)	41.072 (0.399)	39.600 (0.394)	0.136	0.271
Home Index (Family Care Indicators scale)	0.810 (0.020)	0.771 (0.020)	0.791 (0.020)	0.168	0.504
Socio-emotional stimulation (PBC Nurturing Raw scale)	3.995 (0.015)	3.994 (0.014)	4.016 (0.014)	0.967	0.306
Use of disciplinary strategies (PBC Discipline raw scale)	2.729 (0.019)	2.733 (0.019)	2.692 (0.020)	0.877	0.180
Observations	971	971	971		
F-test of joint significance (p-value)				0.599	0.341

Note: T-tests report comparisons between the control arm against NEP-Basico and NEP-Intensivo. Significance levels: *p<=10%, **p<=5%, ***p<=1%. F-test for the joint significance across all variables is reported at the bottom.

Figure A1 indicates that the outcomes of standardized language test at baseline (PLSIV) and of executive functions tests (DCCS) improve as the caregiver’s educational attainment increases. The same is true if we plot the receptive language test TEVIR using endline data. Figure A2 shows similar patterns for socio-emotional development: maladaptive behaviors (internalizing and externalizing behavioral problems) measured through the CBCL decrease as the caregiver’s educational attainment increase, whereas adaptive behaviors, measured using the Battelle socio-personal scale, are positively related with the caregiver’s educational attainment.

Figure A1: Baseline child cognitive development and primary caregiver education

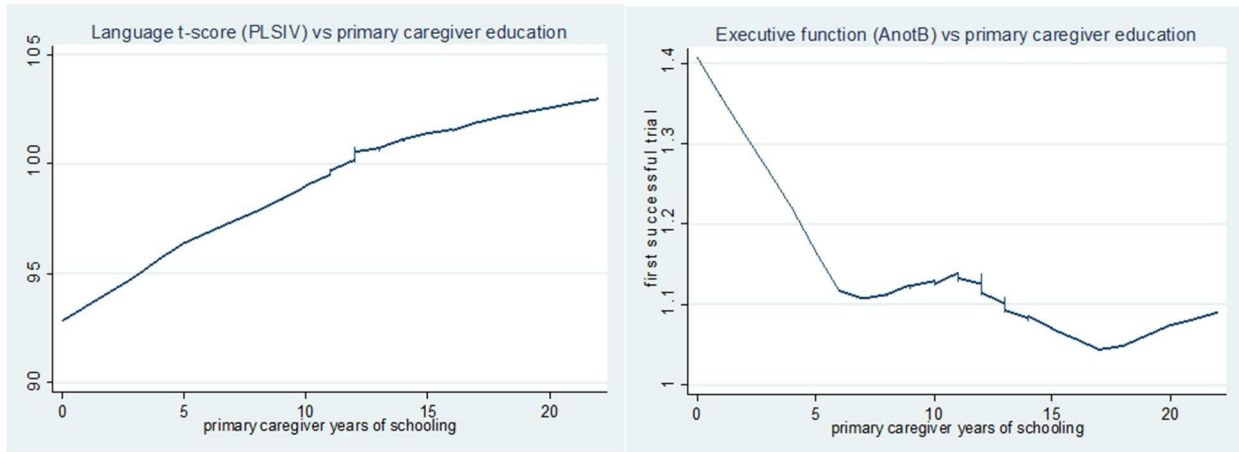


Figure A2: Baseline child behavior and primary caregiver education

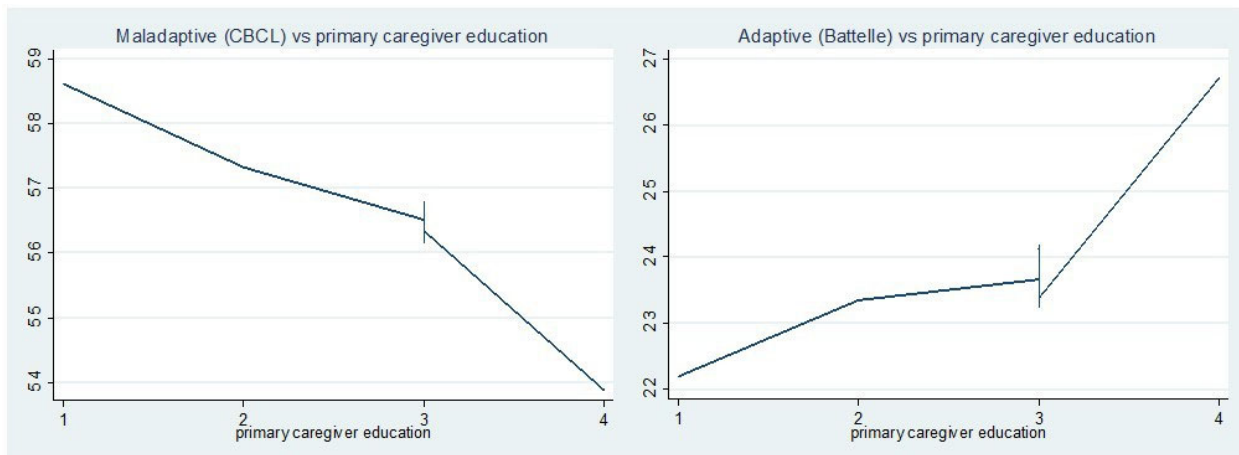


Figure A3 (right side) illustrates that positive cognitive stimulation practices measured through the HOME index are positively associated with the caregiver’s educational attainment. Figure A3 (left side) shows that non-cognitive stimulation practices measured with the PBC nurturing scale also increase with caregiver’s education, while the use of harsh disciplinary practices decreases at higher levels of educational attainment.

Figure A3: Baseline parenting behaviors and primary caregiver education

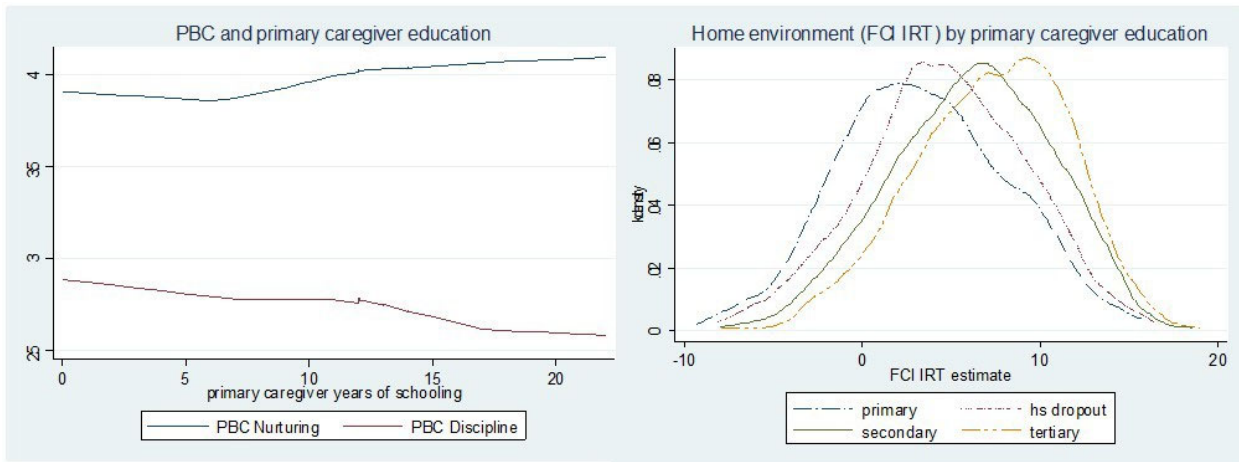
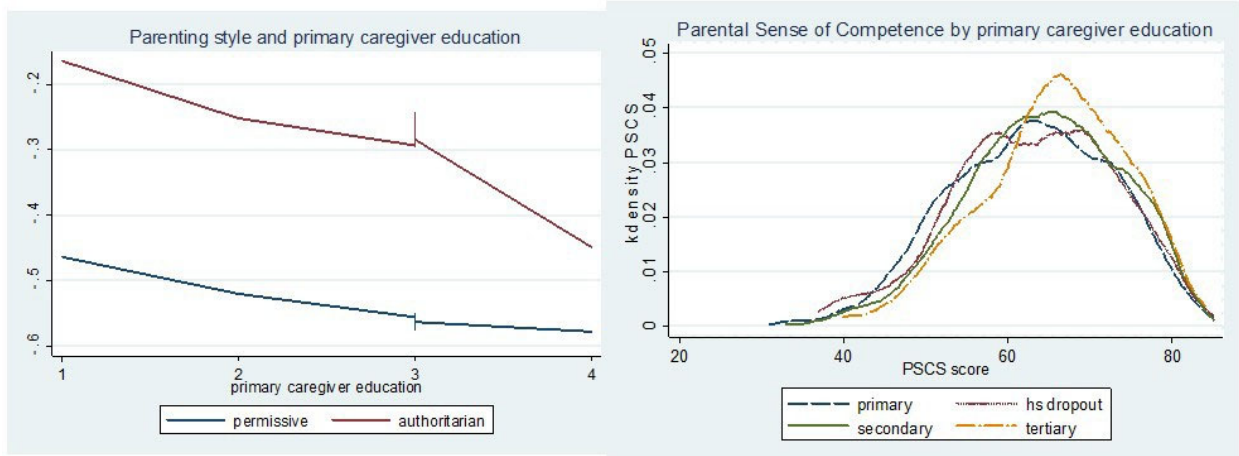


Figure A4 (left side) reveals important socio-economic gradients for parenting styles and parental beliefs. Authoritarian and permissive parenting styles are more present among parents with low educational attainment, in contrast with the authoritative style. Figure A4 (right side) illustrates that both perceived self-efficacy and social support increase as caregiver’s educational attainment increases.

Figure A4: Baseline parental beliefs



Appendix 2: ITT and IV Impacts, sensitivity

Table A6: ITT estimates of child outcomes, with added controls

	Obs.	(i) + (age/gender) and socioeconomic characteristics (caregiver's education, hh'ld income, hh'ld size)		(ii) + maternal IQ and personality traits		(iii) + baseline outcomes	
		NEP B	NEP I	NEP B	NEP I	NEP B	NEP I
Language	2894	0.073 (0.044)	0.098**† (0.045)	0.076* (0.044)	0.103**† (0.044)	0.082* (0.047)	0.115**† (0.047)
Personal-Social Development: Composite Index	1509	0.068 (0.061)	0.127**† (0.061)	0.068 (0.061)	0.127**† (0.061)	0.066 (0.060)	0.126**† (0.061)
Subscale Social Role	2325	0.062 (0.048)	0.094** (0.048)	0.062 (0.048)	0.094** (0.048)	0.062 (0.048)	0.097** (0.048)
Subscale child-adult interactions	1532	0.068 (0.061)	0.137**† (0.062)	0.068 (0.061)	0.137**† (0.062)	0.068 (0.061)	0.131**† (0.062)
Subscale peers- interaction	1521	0.007 (0.063)	0.076 (0.064)	0.007 (0.063)	0.076 (0.064)	0.005 (0.063)	0.075 (0.064)
Behavioral problems: Externalization	1971	-0.032 (0.050)	-0.016 (0.050)	-0.038 (0.049)	-0.035 (0.048)	-0.015 (0.059)	0.002 (0.059)
Behavioral problems: Internalization	1887	-0.044 (0.049)	-0.016 (0.049)	-0.044 (0.047)	-0.030 (0.047)	0.006 (0.058)	0.035 (0.059)
Executive Function Sustained attention	2878	-0.005 (0.044)	0.033 (0.045)	-0.007 (0.044)	0.035 (0.044)	-0.006 (0.044)	0.038 (0.045)

Note: Each line within a column specification refers to a separate regression. All regressions control for health center's fixed effects. Significance levels: *p<=10%, **p<=5% testing individual hypotheses, †p<=10% testing multiple hypotheses.

Table A7: IV estimates of child outcomes, with added controls

	Obs.	(i) +child demographics (age/gender)		(ii) +socioeconomic characteristics (caregiver's education,hh'ld income, hh'ld size)		(iii) + baseline outcomes	
		NEP B	NEP I	NEP B	NEP I	NEP B	NEP I
Executive Function	2879	-0.024 (0.235)	0.121 (0.185)	-0.034 (0.236)	0.128 (0.186)	-0.031 (0.236)	0.14 (0.186)
Language	2895	0.397* (0.241)	0.423**† (0.189)	0.417* (0.242)	0.445**† (0.190)	0.444* (0.254)	0.500**† (0.202)
Personal-Social Development: Composite Index	1509	0.318 (0.305)	0.564**† (0.261)	0.324 (0.297)	0.516**† (0.256)	0.316 (0.294)	0.508**† (0.253)
Subscale Social Role	2325	0.292 (0.248)	0.398** (0.201)	0.316 (0.245)	0.393** (0.200)	0.314 (0.245)	0.403** (0.200)
Subscale child- adult interactions	1532	0.354 (0.312)	0.616**† (0.262)	0.330 (0.304)	0.553**† (0.257)	0.328 (0.301)	0.523**† (0.253)
Subscale peers- interaction	1521	0.008 (0.314)	0.317 (0.267)	0.033 (0.305)	0.289 (0.261)	0.020 (0.304)	0.283 (0.258)
Externalization	1971	-0.173 (0.271)	-0.084 (0.210)	-0.209 (0.263)	-0.161 (0.203)	-0.078 (0.318)	-0.008 (0.256)
Internalization	1887	-0.240 (0.269)	-0.088 (0.202)	-0.242 (0.261)	-0.143 (0.196)	0.038 (0.323)	0.134 (0.248)

Note: Each line within a column specification refers to a separate regression. All regressions control for health center's fixed effects. Significance levels: *p<=10%, **p<=5% testing individual hypotheses, †p<=10% testing multiple hypotheses..

Table A8: ITT estimated parameters of parental practices, with added controls

	Obs.	(i) + child demographics		(ii)+ socioeconomic characteristics	
		NEP-B	NEP-I	NEP-B	NEP-I
Parental Practices					
Home Index	2545	0.052 (0.071)	0.140**† (0.071)	0.077 (0.069)	0.137**† (0.069)
PBC Affection	2545	0.042 (0.046)	0.084* (0.046)	0.052 (0.046)	0.088* (0.046)
PBC Interaction	2545	0.011 (0.049)	0.010 (0.049)	0.019 (0.045)	0.007 (0.045)
PBC Negative discipline	2545	-0.045 (0.047)	-0.080* (0.047)	-0.042 (0.046)	-0.068 (0.046)
PBC Positive discipline	2545	0.054 (0.048)	0.053 (0.048)	0.061 (0.048)	0.052 (0.048)
Parental Beliefs, Attitudes, Perceptions					
Perceived Self-efficacy	2543	0.031 (0.047)	0.098**† (0.047)	0.033 (0.046)	0.085* (0.046)
Perceived Parental Impact on child development					
Perceived Social Support - Family	2545	-0.076 (0.048)	0.003 (0.048)	-0.067 (0.047)	0.011 (0.047)
Perceived Social Support - Friends	2545	0.068 (0.046)	0.080* (0.046)	0.080 (0.045)	0.073 (0.045)
Perceived Social Support - Others	2545	-0.013 (0.047)	0.014 (0.047)	-0.007 (0.047)	0.013 (0.047)
Democratic style	2545	0.030 (0.048)	0.041 (0.048)	0.028 (0.048)	0.032 (0.048)
Authoritarian style	2545	0.032 (0.048)	-0.025 (0.048)	0.029 (0.048)	-0.021 (0.048)
Permissive style	2545	-0.060 (0.048)	-0.014 (0.048)	-0.056 (0.048)	-0.009 (0.048)
Elicited Age High Investment Home Scenario	1487	-0.045 (0.050)	-0.061 (0.049)	-0.056 (0.050)	-0.064 (0.048)
Elicited Age Low Investment Home Scenario	1486	-0.053 (0.050)	-0.103* (0.054)	-0.061 (0.051)	-0.102* (0.054)
Psychological Well Being					
Parental Stress	2545	0.042 (0.047)	-0.011 (0.047)	0.036 (0.046)	-0.006 (0.046)
Depression	2545	0.040 (0.047)	0.034 (0.047)	0.034 (0.046)	0.040 (0.046)

Note: Each line reports estimates from a separate regression. Column (i) adds gender and age, column (ii) adds caregiver's education, hh'ld income, hh'ld size. All regressions control for health center's fixed effects. Significance levels: *p<=10%, **p<=5% testing individual hypotheses, †p<=10% testing multiple hypotheses..

Table A9: IV estimated parameters of parental practices, with added controls.

	Obs.	(i) + child demographics		(ii)+ socioeconomic characteristics	
		NEP-B	NEP-I	NEP-B	NEP-I
Parental Practices					
Home Index	2545	0.429 (0.378)	0.585**† (0.296)	0.441 (0.378)	0.585**† (0.295)
PBC Affection	2545	0.286 (0.251)	0.374* (0.196)	0.310 (0.250)	0.374* (0.195)
PBC Interaction	2545	0.104 (0.248)	0.042 (0.194)	0.093 (0.248)	0.028 (0.193)
PBC Negative discipline	2545	-0.231 (0.254)	-0.292 (0.199)	-0.196 (0.237)	-0.289 (0.185)
PBC Positive discipline	2545	0.333 (0.262)	0.242 (0.205)	0.259 (0.259)	0.194 (0.202)
Parental Beliefs, Attitudes, Perceptions					
Perceived Self-efficacy	2543	0.185 (0.252)	0.352**† (0.198)	0.107 (0.229)	0.325**† (0.179)
Perceived Parental Impact on child development	2545	-0.364 (0.259)	-0.003 (0.202)	-0.342 (0.257)	0.007 (0.201)
Perceived Social Support - Family	2545	0.440* (0.251)	0.337* (0.196)	0.406* (0.245)	0.312 (0.191)
Perceived Social Support - Friends	2545	-0.039 (0.256)	0.044 (0.200)	-0.037 (0.254)	0.038 (0.198)
Perceived Social Support - Others	2545	0.156 (0.260)	0.142 (0.204)	0.057 (0.243)	0.102 (0.190)
Democratic style	2545	0.154 (0.261)	-0.062 (0.205)	0.217 (0.258)	-0.040 (0.201)
Authoritarian style	2545	-0.301 (0.261)	-0.074 (0.204)	-0.269 (0.259)	-0.057 (0.202)
Permissive style	2545	-0.296 (0.256)	-0.249 (0.179)	-0.296 (0.259)	-0.260 (0.180)
Elicited Age High Investment Home Scenario	1487	-0.323 (0.294)	-0.387* (0.205)	-0.329 (0.298)	-0.402* (0.206)
Elicited Age Low Investment Home Scenario	1486				
Psychological Well Being					
Parental Stress	2545	0.196 (0.251)	0.003 (0.197)	0.185 (0.240)	-0.021 (0.188)
Depression	2545	0.187 (0.250)	0.178 (0.196)	0.211 (0.228)	0.171 (0.178)

Note: Each line reports estimates from a separate regression. All regressions control for health center's fixed effects. Significance levels: *p<=10%, **p<=5% testing individual hypotheses, †p<=10% testing multiple hypotheses..

Table A10: First stage (extended table with all controls)

	Participation NEP-B	Participation NEP-I
	Coef (SE)	Coef (SE)
NEP-B	0.200*** (0.012)	
NEP-I		0.262*** (0.013)
Child's age at baseline (base: 0-12 mo.)		
13-24 mo.	0.013 (0.019)	0.027 (0.019)
25-36 mo.	0.024 (0.020)	0.042** (0.021)
37-48 mo.	-0.008 (0.020)	0.027 (0.021)
49-60 mo.	-0.020 (0.023)	0.034 (0.024)
Girls	0.011 (0.012)	0.008 (0.013)
HH Incomes (base: q1)		
q2	-0.006 (0.019)	0.039* (0.020)
q3	-0.005 (0.020)	0.001 (0.020)
q4	-0.023 (0.020)	-0.014 (0.021)
q5	-0.012 (0.022)	0.009 (0.023)
Caregiver Education (base: Primary)		
High School Dropout	-0.007 (0.020)	-0.007 (0.021)
High School Degree	0.031* (0.016)	-0.010 (0.017)
College	0.058** (0.023)	0.037 (0.024)
Number of HH members		
	-0.005 (0.004)	0.001 (0.005)
Single mother	-0.034*** (0.013)	-0.032** (0.013)
Number of younger siblings		
	0.018 (0.013)	-0.008 (0.013)
Caregiver works at baseline		
	-0.030** (0.014)	-0.025* (0.015)
Caregiver works full-time at baseline		
	0.016 (0.019)	0.005 (0.019)
Observations	2530	2530

Note: All regressions control for health center's fixed effects. Significance levels: *p<10%, **p<5%, ***p<1%

Appendix 3: Selective Attrition

In this section we examine the potential importance selective attrition between baseline and follow-up. Table A11 shows that there is some degree of selective attrition that is particularly significant when comparing NEP-B against the control group (Column 1). In the next columns of Table A11 we investigate whether some key outcomes of the study as well as SES variables can explain differential attrition by treatment arm, where we find two results. First, there is a positive and significant Attrition interaction between receptive language at baseline and NEP-B, which is fully explained by a higher language score among non-attrites vs. attrites within the Control Group. And second, there is a negative and significant interaction between the Home Index at baseline and NEP-B, which is fully explained by lower scores among non-attrites vs. attrites in the control group. Interactions between outcomes and NEP-I are never significant.

We correct for potential bias arising from selective attrition using two approaches. First, we adopt a Control Function approach to correct for attrition bias instrumenting attrition with interviewer fixed-effects. The underlying assumptions for the validity of the IV are two: 1) the assignment of interviewers to families is a quasi-random, which we believe it holds as the more than 100 interviewers hired to collect data at baseline were allocated to families according to the municipality of residence of the interviewer; and 2) there is a significant correlation between interviewer and attrition, which we also test to be true. Table A12 shows that the main impacts of the study presented in Table 2 and Table 3 are very similar in magnitude and significance to those that control for attrition.

Second, we further test for potential biases due to selective attrition in our main outcome of interest, receptive language, following Angrist, Bettinger, and Kremer (2006). In their method, they estimate Tobit regressions for a censored outcome for different percentiles of the distribution of the latent variable, assigning the value of the outcome at the percentile for missing values, and the observed outcome for values above the percentile. The idea is to test for the stability of the coefficient of interest across regressions when the percentile is increased, which in our case is useful to test for the robustness of our estimated impacts in language as missing language scores due to attrition are likely to arise from the lower tail of the distribution of the latent outcome, as discussed above. Table A13 presents the Tobit regression outcomes censoring the outcome variable different percentiles and shows that the impacts in language would remain robust regardless of the percentile chosen to censor the data.

Table A11: Attrition: interaction with treatment arms and baseline variables

Dep. Var: Attrition	Treatment	Language	Ex. Function	Behavior	Home	Education	Income
	Coeff/SE	Coeff/SE	Coeff/SE	Coeff/SE	Coeff/SE	Coeff/SE	Coeff/SE
NEP-B	-0.046*** (0.016)	-0.248** (0.098)	-0.108* (0.059)	0.025 (0.094)	0.045*** (0.016)	-0.064** (0.026)	-0.068*** (0.021)
NEP-I	-0.028* (0.016)	-0.163* (0.098)	-0.092 (0.059)	-0.056 (0.092)	-0.029* (0.016)	-0.049* (0.026)	-0.049** (0.021)
Baseline variable		0.002*** (0.001)	-0.011** (0.005)	-0.001 (0.001)	0.015* (0.008)	-0.016 (0.024)	-0.014 (0.023)
NEP-B x baseline variable		0.002** (0.001)	0.007 (0.006)	-0.001 (0.002)	-0.024** (0.011)	0.028 (0.033)	0.050 (0.032)
NEP-I x baseline variable		0.001 (0.001)	0.008 (0.006)	0.001 (0.002)	-0.013 (0.011)	0.033 (0.033)	0.049 (0.033)
Observations	3597	3198	1874	2297	3571	3597	3597

Note: Regressions controls for health center's fixed effects. Significance levels: *p<10%, **p<5%.

Table A12: Control function regression correcting attrition bias

Dep. Var: Final Outcomes	Obs.	NEP-B	NEP_I	P value Test B=I
Executive Function	2879	-0.006 (0.044)	0.038 (0.045)	0.315
Receptive Language	2895	0.080* (0.044)	0.105**† (0.045)	0.602
Personal-Social Development	1509	0.064 (0.061)	0.131**† (0.062)	0.261
Externalization	1971	-0.025 (0.050)	-0.017 (0.050)	0.874
Internalization	1887	-0.030 (0.049)	-0.021 (0.049)	0.849
Home Index	2545	0.083 (0.072)	0.154**† (0.073)	0.320
Self-efficacy	2543	0.035 (0.047)	0.097**† (0.047)	0.179
Perceived parental impact	2545	0.067 (0.046)	0.102**† (0.046)	0.432
Perceived Social Support	2545	0.072 (0.046)	0.083* (0.046)	0.802

Note: Each line within a column specification refers to a separate regression. All regressions control for health center's fixed effects. All regressions include polynomials of fitted values of regressing attrition on interviewer's fixed effects. Significance levels: * $p < 10\%$, ** $p < 5\%$ testing individual hypotheses, † $p < 10\%$ testing multiple hypotheses..

Table A13: Tobit regression for the impacts in language (receptive language at follow-up)

	(1) OLS with observed language coeff/SE	(2) Tobit censored at 2% coeff/SE	(3) Tobit censored at 5% coeff/SE	(4) Tobit censored at 10% coeff/SE	(5) Tobit censored at 15% coeff/SE	(6) Tobit censored at 20% coeff/SE	(7) Tobit censored at 25% coeff/SE
NEP-B	0.074* (0.045)	0.170*** (0.050)	0.185*** (0.054)	0.163*** (0.049)	0.090*** (0.033)	0.076** (0.030)	0.064** (0.027)
NEP-I	0.100** (0.046)	0.133*** (0.050)	0.142*** (0.054)	0.136*** (0.049)	0.079** (0.033)	0.067** (0.030)	0.057** (0.027)
Observations	2895	3576	3576	3576	3576	3576	3576

Note: Each column refers to a separate regression that controls for health center's fixed effects. We adopt the procedure by Angrist, Bettinger, and Kremer (2006), whereby the sample of children with observed language outcomes at endline is censored. Column (1) reports the main impact of NEP without adjusting for censoring. Columns (2)-(7) assume that the data is censored at the 2nd, 5th, 10th, 15th, 20th and 25th percentile and is estimated with a Tobit model. Significance levels: * $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$.

Appendix 4: Construction of child and parental measures

Measurement Error in Outcomes

We first correct for measurement error potentially biasing the standard errors of key outcomes of the study at baseline and follow-up, such as in language development, cognitive stimulation and parental beliefs about childrearing. To do so, we estimated latent factor constructs using standard Item response theory (IRT) methods, which are better suited to predict latent variable scores (the ability or trait) using discrete scale items in psychometric testing (Bock et al. (1996); Lu, Thomas, and Zumbo (2005)). The fundamental building block of IRT is the item characteristic curve (ICC), which links the latent ability, θ , to the probability a randomly drawn examinee of a given ability will answer the item correctly, $P(\theta)$.

For language development at follow-up we estimated a Rasch two-parameter logistic (2PL) for the 116-item receptive language scale TEVI-R, model that is better suited to binary responses. In the 2PL model, $P(\theta)$ varies with ability according to two parameters: a difficulty parameter measuring the item's overall difficulty, and a discrimination parameter, capturing how quickly the likelihood of success changes with respect to ability (Desjardins et al. 2018). Because most responses were incorrect in the last items due to age-characteristics of the sample, convergence was achieved including the 81 first items of the scale.

For home stimulation (HOME inventory) and parental beliefs about child-rearing (Ideas About Parenting scale), we estimated unidimensional models using a partial credit approach, as responses had the form of Likert scales. A partial credit model estimates parameters for the steps within an item (e.g. the parameter for going from a response of 2 to 3 is different from going from response 3 to 4 for a given item) and delivers an estimated continuous attitude score and standard error of measurement on that score, which we use in our regressions. In all the estimations the infit for all of the items in all of the models are reasonable and only few have some evidence of statistically significant misfit.

Age Standardizations

For our measure of receptive language at follow-up (TEVI-R) we have used the standardized scores by age within sample. This approach was adopted after confirming that the original standardized scores would still exhibit age gradients, possibly because our sample is not nationally-representative as the one used by the publishers of the test. Our approach included three steps: a) we estimated regressions of raw scores on cubic polynomials on age and predicted fitted values and residuals; b) we estimated regressions of the variance of the estimated residuals on cubic polynomials on age and used the predicted values to obtain a measure of the age-corrected standard deviation of the variance; c) we used the fitted values obtained in a) and the estimated standard deviation of the variance obtained in b) to standardized raw scores.

A similar approach was used to standardize by age and gender all the relevant measures of parental investments in children, beliefs, and mental health. However, the results are not sensitive to the standardization.

Appendix 5: Heterogeneous Treatment Effects

We examined heterogeneous impacts of NEP on child language and quality of the home environment along two dimensions: socio-economic status/skills and child gender.

Socio-economic status and caregiver/child skills

There exists an increasing interest in interventions and programs with the ability to close early childhood developmental gaps (Heckman 2006). We study differences in the returns of NEP between “advantaged” and “disadvantaged” families by examining treatment effects by caregiver’s education, caregiver’s IQ, and child outcomes at baseline. Table A14 shows heterogeneous impacts by caregiver education. Treatment effects in receptive language among children of lower educated caregivers (high school dropouts or less) are higher than those with more education, although differences across education groups are not significant. In NEP-B, treatment effects are 0.15 SD among the low education group and 0.034 SD among the high education group, and in NEP-I these estimates are 0.12 SD and 0.09 SD, respectively. In socio-emotional outcomes, however, treatment effects are significantly higher (at the 5% level) for the less educated group both in NEP-B (0.25 SD vs -0.03 SD), and in NEP-I (0.3 SD vs. 0.03 SD). The same pattern is observed in caregiver’s behaviors, but differences across groups are not significant. In NEP-B, treatment effects in Home Index is 0.15 SD among the low education group and 0.07 SD among the high education group, and in NEP-I these estimates are 0.22 vs 0.09 SD, respectively.

Table A14: Heterogeneity of impact: caregiver education

Treatment effects by caregiver education	NEP-B			NEP-I		
	Low Educ.	High Educ.	p-value Low=High	Low Educ.	High Educ.	p-value Low=High
Language	0.151** (0.073)	0.034 (0.057)	0.219	0.118 (0.076)	0.087* (0.056)	0.689
Personal-Social Development	0.247** (0.103)	-0.029 (0.077)	0.035	0.303*** (0.107)	0.033 (0.076)	0.042
Home Index	0.149 (0.118)	0.072 (0.091)	0.613	0.251** (0.121)	0.091 (0.089)	0.294

Note: Low education group: caregivers with less than a high school degree education; High education group: caregivers with a high school degree or more. Significance levels: *p<10%, **p<5%, ***p<1%.

Table A15 examines treatment effects by caregiver IQ, measured using the Wechsler Adult Intelligence Scale (WAIS). Caregivers with Low IQ are those below the median of cognitive ability and those with high IQ those above the median. The impacts in language development are significantly larger among the most disadvantaged group. In NEP-B, treatment effects among the Low IQ group are 0.16 SD and 0.00 SD among High IQ group. This difference is statistically significant at the 5% level. In NEP-I, treatment effects among the Low IQ group are 0.20 SD and 0.01 SD among the High IQ group, difference that is also statistically significant at the 5% level. In socio-emotional development, treatment effects are also larger among the Low IQ group, but

this difference is only significant in the NEP-B arm (0.22 SD vs. -0.76 SD). These results are somewhat mirrored by treatment effects by group in the Home Index, but differences between groups are no longer significant.

Table A15: Heterogeneity of impact: caregiver cognition (IQ)

	NEP-B			NEP-I		
	(1) Low: IQ: below median	(2) High: IQ above median	p-value Low=High	(3) Low: IQ below median	(4) High: IQ above median	p-test Low=High
Language	0.160** (0.063)	0.000 (0.064)	0.067	0.197*** (0.064)	0.002 (0.064)	0.032
Personal-Social Development	0.220** (0.089)	-0.076 (0.086)	0.018	0.164* (0.089)	0.109 (0.087)	0.667
Home Index	0.095 (0.104)	0.041 (0.010)	0.920	0.218** (0.104)	0.102 (0.103)	0.438

Note: Caregiver cognition (IQ) is measured with the Wechsler Adult Intelligence Scale (WAIS); Significance levels: *p<10%, **p<5%, ***p<1%.

Finally, Table A16 shows treatment effects in language among above or below the median standardized test score of the PLSIV test of receptive language at baseline. In NEP-B, treatment effects among the Low Baseline Scores group are 0.1 SD and 0.067 among the High Baseline Scores group. In NEP-I, these estimated are 0.16 SD and 0.065 SD, respectively. However, these differences are not statistically significant.

Table A16: Heterogeneity of impact: baseline child outcome

Treatment effects by children's language at baseline	NEP-B			NEP-I		
	(1) Low	(2) High	p-value (1)=(2)	(3) Low	(4) High	p-test (3)=(4)
Language	0.101 (0.067)	0.065 (0.066)	0.707	0.163** (0.068)	0.056 (0.067)	0.278

Note: Significance levels: *p<10%, **p<5%, ***p<1%.

Child Gender

Tables A17 shows treatment effects by child gender, suggesting a larger impact among girls than boys in receptive language. In NEP-B, treatment effects among females are 0.14 SD and among males are 0.02 SD. In NEP-I, treatment effects are 0.13 SD among females and 0.076 among males. However, differences between groups by treatment arm are not statistically significant. We find no significant differences by gender in any treatment arm in socio-emotional development. Moreover, treatment effects by group in the home environment are larger (but not significant) among males.

Table A17: Heterogeneous Treatment Effects by Child Gender

Treatment effects by child gender	NEP-B			NEP-I		
	(1) Male	(2) Female	p-value (1)=(2)	(3) Male	(4) Female	p-value (3)=(4)
Language	0.019 (0.062)	0.135** (0.065)	0.205	0.075 (0.062)	0.120* (0.066)	0.628
Personal-Social Development	0.072 (0.086)	0.063 (0.089)	0.943	0.130 (0.086)	0.148 (0.091)	0.667
Home Index	0.157 (0.103)	0.005 (0.106)	0.306	0.181* (0.101)	0.133 (0.107)	0.752

Note: Significance levels: *p<10%, **p<5%, ***p<1%.