UNITED NATIONS UNIVERSITY
UNU-WIDER

WIDER Working Paper 2020/72

Improving parenting practices for early child development

Experimental evidence from Rwanda

Patricia Justino,[1] Marinella Leone,[2,*] Pierfrancesco Rolla,[3] Monique Abimpaye,[4] Caroline Dusabe,[4] Diane Uwamahoro,[4] and Richard Germond [5]

June 2020

United Nations University World Institute for Development Economics Research
wider.unu.edu

Abstract: This paper investigates the short- and medium-term impact of a randomized group-based early child development programme targeting parents of children aged six to 24 months in a poor, rural district of Rwanda. The programme engaged parents through sessions that included a radio show and facilitated discussions during 17 weekly village-level meetings. Twelve months after baseline, children's communication, problem-solving, and personal social skills improved in treated groups. After almost three years, the effects on child development outcomes in the full treatment group persisted (by 0.18 SD). The intervention also resulted in increases in maternal time investments in both the short and the medium term, as well as improvements in parents' perceptions about their self-efficacy and influence over their child's development. A linear mediation analysis shows that 20 per cent of the positive changes observed in child development can be attributed to the increase in time mothers spent engaging with their children.

Key words: early child development, parenting programme, radio, randomized controlled trial, Rwanda

JEL classification: I15, I25, J13, O15

Acknowledgements: We acknowledge funding from the British Academy Early Childhood Development Programme, supported under the UK Governments Global Challenges Research Fund and by the UK Department for International Development. We thank the participants in the ECD workshop in Kigali in June 2019 as part of the National Early Childhood Development Program Conference. We are also grateful for useful comments from the participants in several seminars and conferences: the NEUDC conference at Northwestern University, UKFEIT conference, CIES conference, Bocconi University Food for Thought seminar, and seminar at UNICEF Innocenti Office of Research. Data were collected after ethical clearance by the Institute of Development Studies Ethics Committee and the Rwandan National Ethics Committee. We are indebted to these institutions for inputs, corrections, and final approval. We thank also all Save the Children Rwanda staff for administrative and logistical support. We are particularly grateful to all parents and children who participated in the study, the local facilitators that ran the parenting sessions, and the enumerators that collected the data. All mistakes are our own. The views expressed herein are those of the authors, and do not necessarily reflect the views of the British Academy and of the UK Department for International Development.

Notes: tables and figures at the end.

[1] Institute of Development Studies, University of Sussex, Brighton, UK, and UNU-WIDER, Helsinki, Finland; [2] University of Pavia, Italy, and Institute of Development Studies, United Kingdom; [3] Institute of Development Studies, United Kingdom; [4] Save the Children International, Kigali, Rwanda; [5] Save the Children UK, London. *Corresponding author: marinella.leone@unipv.it.

This study is published within the UNU-WIDER project Social mobility in the Global South – concepts, measures, and determinants.

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9256-829-0

https://doi.org/10.35188/UNU-WIDER/2020/829-0

Typescript prepared by Siméon Rapin.

The United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

1 Introduction

An estimated 279 million children from low- and middle-income countries are at risk of not reaching their development potential due to extreme poverty and stunting (Grantham-McGregor et al. 2007). Children from poor families and those exposed to risk factors tend to accumulate developmental deficits from a very early age (Heckman 2006; Cunha and Heckman 2007; Grantham-McGregor et al. 2007; Lu et al. 2016; Black et al. 2017). For instance, in Rwanda, only seven per cent of children are developmentally on track in areas of literacy and numeracy (NISR, MOH, and ICF International 2015b). Literacy levels among parents in poor countries are very low. In Rwanda, only 20 per cent of adult women and 25 per cent of adult men are literate (NISR, MOH, and ICF International 2015b). Under these circumstances, even though parents might be aware of the importance of early child development investments, parenting practices may not always be conducive to the positive development and stimulation of young children.[1] Parents in economically disadvantaged families also face serious financial constraints, and experience several stress factors that may impact negatively on their engagement with their children. Consequently, this shapes how their children's cognitive and non-cognitive skills develop throughout childhood to achieve their full potential. These deficits in early life may be transmitted across generations, leading to a perpetuation of poverty and low child development outcomes over time (Doepke et al. 2019).

Therefore, early childhood interventions may be particularly valuable at improving child development outcomes when they target disadvantaged families (Heckman 2006; Walker et al. 2011; Gertler et al. 2014). However, much of the existing evidence draws on interventions implemented in the USA, on programmes implemented among poor families in middle-income developing countries that already have a well-functioning welfare system in place, such as Colombia and Chile (Andrew et al. 2018; Carneiro et al. 2019a), or among disadvantaged families in developed countries, such as Ireland (Doyle 2020). Evidence of the effects of such interventions amongst the poorest of the poor is scarce,[2] and rural, remote areas are rarely covered by such programmes.

This paper addresses this gap in the literature by evaluating an early child development programme implemented among some of the poorest rural communities in the world. To the best of our knowledge, this is one of the first times that such a programme is implemented and evaluated using a randomized trial among highly deprived families living in remote rural locations—and the first in Rwanda.

A large body of research shows that the first years of life are crucial for lifelong outcomes. Brain development is particularly rapid and malleable during that period (Knudsen 2004), and investments in children's development at a young age determine their future human capital accumulation, earnings, and health status (Currie and Almond 2011; Almond et al. 2018). The cognitive and non-cognitive skills of young children are shaped by a variety of factors, including the economic and social status of their families (Heckman and Carneiro 2003; Attanasio et al. 2020), parenting styles (Doepke and Zilibotti 2017), the neighbourhood in which they live (Katz et al. 2001; Chetty et al. 2016), and macroeconomic factors, such as inequality and poverty constraints (Heckman and Carneiro 2003; Doepke et al. 2019). As a result, early child development interventions take a variety of forms, ranging from center-based preschool interventions (Aboud and Yousafzai 2015; Ozler et al. 2018; Andrew et al. 2019) to home visits by trained practitioners, sometimes combined with psychosocial stimulation and micronutrient supplementation (Grantham-McGregor et al. 1991; Walker et al. 2005; Campbell et al. 2014; Kazianga et al. 2014; Andrew et al. 2018; Doyle 2020). Increasingly, early childhood programmes target parents directly,

[1] Many children in Rwanda are subjected to violent forms of discipline and a third receives inadequate care, being left either alone or in the care of another child under the age of ten (NISR, MOH, and ICF International 2015b).

[2] One exception is the widely studied Jamaica home visiting programme (Grantham-McGregor et al. 1991), which targeted children with particularly large initial disadvantages, such as malnutrition and low cognitive skills.

generally through home visits and, more recently, through group sessions. These programmes aim to support positive parenting practices and stimulation (Britto et al. 2017; Attanasio et al. 2020), recognizing the central role of parents and their behaviour in shaping child development outcomes (Todd and Wolpin 2007; Dooley and Stewart 2007; Cunha et al. 2013). In general, evaluations of such programmes have shown that improving parenting practices and behaviours are key for the success and long-term sustainability of early childhood interventions (Attanasio et al. 2015, 2020; Britto et al. 2017).

This paper follows in this tradition, and investigates the short- and medium-term impact of an early child development intervention that targeted parents of children aged six to 24 months in rural Rwanda, called First Steps (*Intera za Mbere*). The programme was developed and implemented by Save the Children in collaboration with Umuhuza, a Rwandan NGO, in the (rural) Ngororero district in the Western Province of Rwanda. Ngororero is one of the poorest and most remote areas of Rwanda—itself one of the poorest countries in the world—a context seldom covered by early child development interventions.[3]

First Steps is a group-based intervention that trained parents through 17 weekly group meetings. During each meeting, parents listened to a radio drama, where each episode focused on a key parenting practice. The radio drama was both preceded and followed by group discussions with trained facilitators. The aim of these group sessions was to improve the quality of parent–child interactions by equipping parents with the skills to engage more closely with their children in developmentally-appropriate learning activities, centered around daily routines and using household resources as learning tools. Each group session followed a curriculum defined beforehand, focusing on a specific topic related to child development.[4]

The programme was designed as a cluster-randomized controlled trial with three treatment arms. In the first arm—the control group—parents did not receive the intervention. In the second arm—the light treatment (LT) group—parents attended 17 weekly meetings, which included a group radio listening component and discussions led by a facilitator who received training and an activity booklet, which outlined the activities, games, and key messages to share with parents in each session. The third arm—the full treatment (FT) group—received further inputs in addition to the components of the light treatment group: a supervising facilitator recruited at the cell level to support the village facilitator, one home visit, take home leaflets that reinforced the main messages discussed during the group sessions, and a children's book gifted to each family.

We collected baseline, end-line, and follow-up data in August 2015, September 2016, and May 2018, respectively. The intervention started in November 2015, and ended in April 2016. In this paper, we investigate the short-term (12-months) and medium-term (33-months) impact of First Steps on key dimensions of child development and parenting practices. In line with most of the existing literature on early child development programmes, this study analyses the impact of the programme on a rich set of child development outcomes. These include communication ability, gross motor skills, fine motor skills, problem solving ability, and personal social interactions. In addition, the design of the study allows us to evaluate the impact of the intervention in terms of changes in parental behaviours. We do so by examining the direct impact of the intervention in the short and medium terms on two key parenting variables: parental (mother and father) time investments in interactions with the child, and parental (mother and father) perceptions of their ability to influence their children's development trajectories (parental self-

[3] Rwanda is listed 158th in the Human Development Index list (UNDP 2018) and listed 146th in terms of GDP per capita (IMF 2019). The Ngororero district is one of the poorest districts in Rwanda with around 48 per cent of the population under the poverty line and with 21.5 per cent of its population in extreme poverty (NISR 2018b). In the West Province of Rwanda, 45 per cent of children under five years old are stunted (56 per cent in Ngororero district) compared to 38 per cent at the national level (NISR, MOH, and ICF International 2015a).

[4] The following topics were included: child development, nutrition and breastfeeding, health, positive discipline, early literacy, early math, responsive caring, and play.

efficacy). At follow-up, we were also able to analyse the effect of the programme on attitudes towards gender norms, on locus of control, and on aspirations. Parental investment and parental perceptions and beliefs are at the core of children's human capital production function, particularly during their early stages of life (Attanasio et al. 2020). Recent early child development programmes similar to First Steps suggest that these parental outcomes represent key pathways to sustained child development improvements (Carneiro et al. 2019a). Moreover, increasing skills and knowledge may only lead to behavioural change if parents also have confidence in their abilities. Parental self-efficacy (PSE), in particular, has been shown to be a predictor of parenting skills outcomes in intervention programmes (Spoth et al. 1995) as well as a predictor of child treatment response outcomes (Hoza et al. 2000; Carneiro et al. 2019a).

We estimate in the paper the effect of First Steps on parental time investment and parental beliefs, and then conduct a mediation analysis of factors that may explain the impact of the programme on child development outcomes across different time periods (Kim et al. 2018; Carneiro et al. 2019a). We decompose linearly the treatment effects on child outcomes into components attributable to changes in parental investments, and in how parents perceive their influence over their children's development.

The results of the study show that the First Steps programme had a positive and substantial impact on child development outcomes after 12 months. Children whose parents participated in First Steps experienced an increase of 0.3 (LT) and 0.4 (FT) standard deviations (SD) in the aggregate average index of child development outcomes, relative to children in the control group. These effect sizes are large and consistent with recent studies conducted in more advanced economies (Carneiro et al. 2019a; Doyle 2020). We also find large 12-month impacts of the programme on parental time investments and on maternal perceived influence over their child's development. These results show that the effects of the programme on the aggregate average index of maternal time investments and maternal perceived influence range between 0.4 SD (LT group) and 0.6 SD (FT group).

After almost three years, the full treatment arm still shows large effects on child development outcomes (0.2 SD) and on the aggregate average index of maternal time investment (0.2 SD). We also find an increase of 0.13 SD in the aggregate mean index of maternal self-efficacy for those who received the full treatment as well as positive impacts on attitudes towards gender roles, on locus of control, and on aspirations. The programme appears thus to have had large direct effects on child development, as well as on behavioural changes related to mothers' engagement with their child and levels of confidence in own abilities. The linear mediation analysis shows that maternal time investments yield important returns in terms of development outcomes in the short and longer terms and account for about 20 per cent of the impact of the First Steps programme on child development outcomes. These changes in parental behaviour thus play an important role in ensuring the persistence of the positive effects of the intervention over the longer term. All results are robust to tests accounting for multiple hypothesis testing, baseline differences, selection bias due to attrition, and social desirability biases and to an alternative construction of the main outcomes.

This paper entails important contributions to the existing literature on parenting programmes and, more generally, on early child development interventions. First, the paper adds to a limited set of studies that investigate the impact of parenting programmes in poor rural areas of sub-Saharan Africa, and is the first in Rwanda.[5] We find substantial impacts of the programme, suggesting that early child development interventions can produce effects—both on child development outcomes and on parental behavioural changes—that persist across time amongst some of the most deprived communities in the world.

[5] To our knowledge, two studies have evaluated group-based parenting programmes in poor settings of developing countries, such as Madagascar and Sierra Leone (Fernald et al. 2019; Chandra et al. 2020), but are more limited in scope than First Steps. Carneiro et al. (2019b) evaluate the combined impact of a cash transfer programme that targets pregnant mothers and of information provided to parents on best practices related to pregnancy and early life nutrition in Northern Nigeria.

Second, the study is based on data collected in two periods—12 months and 33 months after the intervention—which provides us with a rich dataset to analyse changes in the programme impact across time. Well-known interventions in the early child development literature have shown mixed long-term evidence on child development and the home environment (Grantham-McGregor et al. 1991; Campbell and Ramey 1994; Walker et al. 2005; Heckman et al. 2010, 2013; Gertler et al. 2014; Bailey et al. 2017). Some studies show positive effects of early child development interventions over time (see review in Doyle 2020), whilst others report substantial fading out effects in the longer term (Bailey et al. 2017). The closest paper to our study is Carneiro et al. (2019a). It analyses the medium-term impact of a large-scale early child development programme implemented in Chile, but is not able to compare these with shorter-term effects.

Third, this study entails important contributions to the literature on the design and implementation of early child development interventions, which are related to the nature of the First Steps curriculum. Despite growing evidence on the effectiveness of early child development programmes (Britto et al. 2017; Doyle 2020), there is still limited knowledge about which curriculum is best for scaling-up such programmes, and ensure their impacts are sustainable over time (Carneiro et al. 2019a). There is some discussion in the literature over what interventions may work best to improve development outcomes among young children, particularly with respect to the advantages and disadvantages of group sessions in relation to (more expensive) home visits. The literature on the effects of group-based interventions is somewhat mixed. In general, group-based interventions have been shown to change parental knowledge and promote positive parenting behaviours. However, not all interventions systematically improve child development outcomes (Engle et al. 2011; Black et al. 2017; Britto et al. 2017). Group-based programmes are often less expensive than home visits, as they encourage peer-to-peer learning and support, and have the potential to modify group norms with respect to child raising and education (Aboud and Yousafzai 2015; Carneiro et al. 2019a). Group sessions combined with some home visits seem to be the most promising model of parenting stimulation programmes, but the evidence is still scarce.

First Steps was designed with a combination of different components, some of which are quite novel and provide good potential for further analysis. In particular, First Steps is, to our knowledge, the first programme in which group meetings included a live radio listening component seamlessly weaved into the core meeting activities and built around the curriculum. Although we are unable to isolate the effect of the radio programme on its own, the mounting positive effects on the effectiveness of radio and other media to promote social change and development (Paluck 2009; Bernard et al. 2015; La Ferrara 2016; Banerjee et al. 2019) suggest that the radio component may have contributed significantly to the large programme impact we observe in both the short and medium terms.

Another promising component of First Steps, present in only a few studied interventions (e.g., Attanasio et al. 2020), is that it was delivered by facilitators trained by the programme and drawn from the local community with no specific prior experience or expertise. This dimension of the programme has two potential advantages. First, the selected facilitators might be known by the beneficiary families and, therefore, may be more trusted in the community. Second, the costs for the facilitators were relatively low, which may make programmes like First Steps more feasible in poorer countries. For instance, the facilitators received an incentive which was one tenth of the stipend given to the workers in the renowned Lady Health Workers Programme in Pakistan (Yousafzai et al. 2014), and yet the positive effects of First Steps are large and sustained over time.

Given the context in which it was being implemented, First Steps also focused on improving the quality of daily interactions between parents and their child based on activities that could be conducted alongside daily routines. For instance, parents were taught to talk to children while cooking, when pre-primary school was closed, or while working in the field. Parents in the villages covered by First Steps were also incentivized to coordinate the daycare of children in turns, and were taught how to utilize household resources as learning tools (for instance, parents learned how to make a book with the materials found

at home and in the surroundings). This emphasis on improving engagement between parents and child alongside daily routines may have enhanced parenting skills without adding to the daily stress of coping with low incomes in very remote areas. These small nudges may well be applicable to many other similar low-income contexts where parents are not able to take time from subsistence activities to engage more productively with their children. In addition, the full treatment group in First Steps included the provision of a children's book that was gifted to families as well as take-home cards, which included illustrations and texts with key messages introduced in each session. These additional components could have in principle relieved a constraint where, on average, families in very poor and deprived communities have almost no children's books in the house and very limited literacy materials.[6] The stronger effects observed in the full treatment group suggest that these additional components played an important role, even though we are not able to disentangle their individual effects.

Finally, by design, the time investments of parents into the First Steps programme was of moderate duration. First Steps was as intense and frequent as other early child development programmes (two hours, once a week), but its duration (17 weeks) was lower than other widely studied parenting programmes (Britto et al. 2015), suggesting that low intensity programmes—perhaps more easily implemented in remote rural areas—can also have substantial impacts on child development outcomes among economically disadvantaged populations. In fact, the impacts we document of First Steps are comparable in magnitude to more intensive and expensive interventions based on home visits, group-based interventions, or a combination of both implemented elsewhere in the world (see review in Doyle 2020).

The remainder of the paper proceeds as follows. Section 2 provides the background of the study and a description of the programme, its design and sampling strategy. Section 3 describes the data collected and the main outcomes of interest. Section 4 discusses the empirical strategy, and Section 5 presents the main findings. Section 6 shows results from heterogeneous treatment effects. Section 7 provides a discussion of the potential mechanisms and results from the mediation analysis. Section 8 concludes the paper.

2 Background and experimental design

2.1 The First Steps programme

First Steps is a group-based early child development programme that targeted parents of children aged between six and 24 months old.[7] It is a participatory programme in which parents, along with their children, were invited to attend 17 weekly group meetings that included listening to a radio drama and participating in facilitated discussions.[8] During the weekly meetings, parents met in a central location in their village (e.g., primary school, village leader office, or even outdoor) to reflect about the previous week's session, listen to a new episode of a radio drama, discuss its content with a local fa-

[6] According to estimates in DHS Rwanda, on average, only one per cent of families with children under five years have books at home (NISR, MOH, and ICF International 2015b). This is worrisome as the presence of books at home is a strong predictor of cognitive and language skills (Tomopoulos et al. 2006).

[7] First Steps was initially designed to target parents of children aged between 0 and 36 months. During the implementation it was decided to engage only parents of children aged between six and 24 months. The only exclusion criteria for the purposes of the research project was children with disabilities as none of the intervention arms was specifically designed to target children with disabilities. An early discussion of the intervention is done in Abimpaye et al. (2019). The term 'parent' refers to the person who is most involved in raising the child (i.e., the principal caregiver of the child). At baseline, around 93 per cent of principal caregivers were mothers and five per cent were fathers. The remaining two per cent were grandparents or other family members.

[8] Both parents were invited to participate in the weekly meetings. The presence of the principal caregiver of the child was required, but the other parent was also encouraged to attend.

cilitator, and learn simple, age-appropriate activities and games they could use at home to support their child's development. The radio drama was developed by Save the Children and Umuhuza. The seventeen episodes depicted a fictional community in which a parenting programme was being implemented. Once the radio episode was aired (each episode lasting 15–20 minutes), the facilitator discussed the radio episode with the parents for about 30 minutes to one hour. The plot followed the change experienced by the characters as they addressed various parenting practices, attitudes, and beliefs, including the role of fathers in childcare and development. The radio programme was both preceded and followed by a participatory conversation between the village facilitator and the parents. The meetings also involved parents practicing games and activities with their children. The aim of the group sessions was to improve the quality of parent–child interactions and to equip parents to engage with their children in developmentally-appropriate learning activities, centered around daily routines, and using household resources as learning tools. It also aimed to support parents with knowledge about feeding, nutrition, and child health. The content of these sessions included (i) early communication and promotion of emergent literacy at home; (ii) learning through play; (iii) responsive care and bonding; (iv) nutrition and health.[9] To achieve these objectives, the discussions with parents included the use of posters, which highlighted key parenting practices and messages, and involved practicing activities and games between parents and children. In addition, parents were trained in how to make homemade books and toys using locally available materials.

The village facilitators were drawn from a network of local women and men. Facilitators received training during 3.5 days and were paid RWF4,000 per month as an incentive (approximately US$5). All village facilitators also received an activity booklet, which outlined the activities, games, and key messages to share with parents in each session. In some villages (which received the full treatment), one additional supervising facilitator was recruited at the cell level to support the village facilitators in sessions and home visits. These facilitators were also trained during 3.5 days, and received a slightly higher pay than the other facilitators (RWF4,500 per month).[10]

2.2 Study design and sample selection

The intervention was evaluated using a cluster-randomized controlled trial with a control group and two treatment arms. Within the Ngororero district, nine sectors (out of 13) were selected for the study. Appendix A.2 provides further details about the sample selection and intervention areas. The intervention was randomly assigned at the sector level to three groups, composed of 27 villages each: a control group, a light treatment (LT) group, and a full treatment (FT) group.

The first arm, the control group, did not receive any treatment. Parents in control group villages were invited to participate in the programme, but were told that it would be offered at a later date.[11] Both LT and FT groups were offered group-based parenting sessions supported by a village facilitator, as described above. In addition, the FT group also received: (i) additional inputs from a supervising facilitator recruited at the cell level to support the village facilitators; (ii) one home visit by the village and cell-based facilitators; (iii) provision of leaflets for parents to take home after each session reminding them of each session contents; (iv) a child's book gifted to each household upon completion of all sessions. Figure 1 summarizes the experimental design.

---

[9] Appendix A.1 lists the specific sessions covered under each main topic.

[10] Appendix A.1 provides more details about the recruitment of facilitators.

[11] The control group was supposed to receive the programme in October 2016 after the end-line data were collected in September 2016. However, only 14 out of 27 villages in the control group received the programme; and those who received it experienced serious implementation issues. As a robustness check, we investigated whether the 33-month effects change by excluding from the control group villages that received the intervention after the end-line. Results are mostly unchanged and available upon request.

The programme was randomized at the sector level because the weekly meetings included the live radio component. Therefore, by implementing the intervention arms at the sector level, risks of contamination were reduced as control villages were less likely to hear about the First Steps radio programme from sector officials or from First Steps participants in community events. These villages did not receive information about the group meetings or the radio show air dates and time, and a not so popular radio station was selected to reduce the likelihood that the control group would accidentally listen to the programme.

In all selected villages—whether treatment or control—all families with a child aged six months to 24 months were eligible to participate in the programme, and all eligible families were invited to participate. In each village, if there were more than 20 eligible families, 20 families were randomly selected for the study.[12]

Participation of parents in the First Steps meetings was voluntary. Almost all families who were offered participation in the study accepted and participated in at least one session.[13] Compliance rates, calculated as the ratio between the number of participants at end-line and the number of people assigned to the treatment at baseline, was 85 per cent for the FT group and 89 per cent for the LT group.

3 Data and measurement

The First Steps intervention was conducted between November 2015 and April 2016. Baseline data were collected in August 2015. End-line data were collected in September 2016. We collected follow-up data in May 2018 in order to understand whether and how initial effects may have persisted across time. An illustrative depiction of the timeline is presented in Figure 2. At baseline, the study included an average of 540 children in each intervention arm, resulting in a total sample of 1,614 children and their parents. In September 2016, due to an average attrition rate of ten per cent, the total sample included 1,452 children and their parents. In May 2018, we were able to track and interview 1,320 parents and 1,278 children.[14] We tested for differential attrition in group assignment and baseline characteristics. When controlling for baseline variables, we did not find evidence of differential attrition bias (discussion and results in Appendix C.1).

3.1 Child development measures

We collected data to measure child development outcomes using an adapted and translated (into Kinyarwanda) version of the Ages & Stages Questionnaires 3d edition (ASQ henceforth). The ASQ is a well-established child development screening tool, and is implemented without the need for additional professionals.[15]

The ASQ questionnaire requires the principal caregiver to report on five child development domains. For certain activities, the caregiver is required to interact with the child, while the enumerator observes and

[12] If the number of interested families per village exceeded 20, the remaining interested families were placed on a waiting list, and told that they would receive the programme at a later stage.

[13] On average, parents attended 12 sessions out of a total of 17 sessions.

[14] The difference between the children and parents sample size is due to the fact that, during the follow-up survey, we were unable to interview 42 children that were away from home at the time of the interview, and were only able to interview their parents.

[15] Overall, the literature suggests that there is considerable agreement between the ASQ and standardized measures that are conducted by professionals (Skellern et al. 2001; Gollenberg et al. 2010; Doyle et al. 2017), such as the Bayley Scales of Infant Development (Bayley 2006).

records the information from a distance.[16] These include activities in the domains of communication, gross motor skills, fine motor, problem solving, and half of the questions on personal social skills. Other questions are self-reported by the caregiver. Because self-reported answers may be subject to potential social desirability bias, we have conducted a number of tests to assess the potential extent of such a bias. Tests are reported in Appendix C.2. We find no evidence that social desirability bias may affect the results of the programme evaluation.

Child development is assessed by means of 30 questions asked in relation to six activities intended to capture five distinct child development dimensions: (i) communication; (ii) gross motor skills; (iii) fine motor skills; (iv) problem solving ability; and (v) personal-social interactions. Given the age-specific nature of the questions, the questionnaire was specifically tailored to different age ranges. At the end of the assessment, a total score was produced for each skill by adding up all the activity scores. We calculated standardized scores for each child development dimension with respect to the control group in the relevant survey wave (by subtracting the mean and dividing by the standard deviation of the control group). We obtained five standardized scores for each child development dimension. Appendix A.3 provides detailed information about the ASQ tool and the construction of the main outcome variables. To draw general conclusions about the experiment's results, we aggregated the five indicators into one mean index defined as 'child development aggregate index', which aggregates information over multiple treatment effect estimates (Kling et al. 2007). This index is defined to be the equally weighted average of z-scores of its components, with the sign of each measure oriented so that more beneficial outcomes have a higher score. We calculated the index by averaging the z-scores of the five child development dimensions, following Kling et al. (2007). We discuss a correction for multiple hypothesis testing in Section 4.

3.2 Household and principal caregiver measures

As discussed in the introduction, in addition to child development outcomes, we also collected detailed data on the home environment, parenting practices, and parental behaviour. These data were collected using the Home Observation for Measurement of the Environment - Short Form (HOME-SF) tool adapted for the Rwandan context. Appendix A.3 provides more details about this tool and the variables we constructed. Importantly, the questionnaire includes questions about primary caregivers' time investments in activities they engage with their child, which has allowed us to investigate how parental engagement may mediate the impact of the programme on child development outcomes. Information collected includes the frequency of interactions across a set of activities the primary caregiver performed with her child and similar practices followed by her/his partner. In most cases (93 per cent), these questions were answered by the mother, who answered questions for herself and also on behalf of her husband. In five per cent of the cases, the principal caregiver was the father. In approximately two per cent of the cases, the principal caregiver was neither the mother nor the father.[17]

The caregiver was asked to report about the frequency of interactions with the child across eighteen activities, including (i) positive discipline activities, such as praising, appreciation, and soothing when the child is upset; (ii) learning/play activities, such as playing, singing, and reading picture books; and (iii) negative discipline activities, such as criticizing, threatening, hitting, pushing, and spanking the child. For each activity, we created a standardized score by subtracting the control group mean and dividing it by the control group standard deviation of the relevant survey wave. We constructed an aggregate index by taking the average of the standardized scores. The resulting indicators are defined as

[16] For instance, in one example among many, in the ASQ questionnaire for children of six months in the gross motor section (question 1), the enumerator asks: 'When you put your baby on the floor, does she lean on her hands while sitting?'. The caregiver proceeds in the activity while the enumerator observes from a distance and reports the answers.

[17] If for example the principal caregiver is the grandmother, she was asked to report also about the activities performed by her spouse (the grandfather).

maternal time investment and paternal time investment. Maternal (paternal) time investment measures correspond to observations related to the mother (father) when she (he) is the respondent.[18] As discussed in more detail in Appendix A.3, in the follow-up survey conducted in May 2018, we collected data on 15 activities, because we grouped together similar activities. Due to time and budget constraints, the respondent was asked to report about each activity using a binary indicator rather than the frequency measure.

As part of the HOME-SF questionnaire, we also recorded information about parental influence across six dimensions of their child's lives: (i) child's learning, (ii) child development, (iii) nutrition, (iv) child care, (v) discipline or child guidance, and (vi) health care. As above, we calculated standardized scores with respect to the control group in the relevant survey wave, and constructed an aggregate index for each standardized outcome variable. We report below results for the six domains and the average aggregate index. We distinguish between maternal influence (when the respondent is the mother) and paternal influence (when the respondent is the father). We did not collect this outcome separately at follow-up, because we collected detailed data on parental self-efficacy, which include this parental influence dimension.[19]

At follow-up, we collected parental self-efficacy measures administering the *Tool to Measure Parenting Self-Efficacy* (TOPSE) (Kendall and Bloomfield 2005). In this questionnaire, the respondent was asked to provide answers about self-efficacy statements using a scale from 1 (disagree a lot) to 5 (agree a lot) across eight different dimensions: (i) emotion and affection; (ii) play and enjoyment, (iii) empathy and understanding; (iv) control; (v) discipline and setting boundaries; (vi) pressures; (vii) self-acceptance; and (viii) learning and knowledge.

We expanded the HOME-SF at follow-up by asking questions to the caregiver about additional dimensions, including parental attitudes about child health, behaviour, development, and gender roles in the household; locus of control; and caregiver's aspirations for themselves and their children. We calculated standardized scores for these outcomes as above.[20]

3.3 Descriptives

Baseline balancing

We examined whether observable baseline characteristics were balanced among treatment arms. Column 1 of Table B1[21] shows the averages for control group characteristics. Columns 2 and 6 show the mean differences between characteristics of LT, FT, and control groups, respectively. The average child development raw score based on the ASQ is 40 at baseline (out of a maximum score of 60) and largely similar across the five dimensions. P-values in columns 3 and 4 show that some child characteristics in the LT group are not fully balanced with respect to the control group mean at baseline. In particular, the LT group includes more girls than the control group, although the difference is not large. Mothers in the LT group seem to be more educated than those in the control group. The LT group also displays higher ASQ scores at baseline than the control group. In Appendix C.3, we provide results from empir-

[18] The number of respondents corresponding to fathers is very small. As discussed above, each respondent is also asked to report on her partner time investment but we preferred not to use these reported measures as our main estimates. However, results using these measures do not change and are available upon request.

[19] In a separate paper (Justino et al. unpublished), we study the causal impact of an intervention that aims at boosting parental self-efficacy, and investigate its effects on self-efficacy and parents' practices.

[20] Questions about these indicators are only available in the follow-up datasets and were not asked in the baseline or end-line surveys.

[21] Tables starting with a 'B' are accessible in Appendix B.

ical checks that account for the observed imbalance in some of the baseline characteristics. All results are robust to the tests performed, which reassures that these imbalances are not likely to affect the final results.

Programme participation

On average parents attended 12 sessions out of a total of 17 sessions. At end-line and follow-up, respondents reported which sessions they remembered having attended (Table B2). Seventy-one per cent of parents reported their participation in the child development's session, and more than 50 per cent of parents reported their participation in the responsive caring and play sessions. Reported participation in the early math session was low (18 per cent on average). As expected, programme's attendance intensity is positively correlated with mothers' education and household wealth (Table B3). As discussed, both parents were invited to participate. On average, 86 per cent of mothers attended First Steps sessions alone. Ten per cent of mothers participated in First Steps jointly with their husband, and only 1.2 per cent of fathers participated alone.

4 Empirical strategy

The randomized nature of the First Steps intervention allows us to identify the causal impact of the programme on child development and parenting outcomes. To that purpose, we estimated the following model for each survey round:

$$y_{ijt} = \alpha + \beta^L T_j^L + \beta^F T_j^F + \lambda y_{ij0} + \gamma X_{ij0} + e_{ijt} \quad (1)$$

where y_{ijt} is the outcome for individual i, in sector j surveyed at time t. t is equal to 0 for baseline, to 1 for end-line and to 2 for follow-up observations. We estimated equation (1) for each round separately. The terms T_j^L and T_j^F are binary indicators for LT and FT treatment sector-level interventions. y_{ij0} is the baseline level of the outcome for individual i in sector j, and X_{ij0} are baseline characteristics. The regressions control for child age and gender, number of children in the household, the primary caregiver age, binary indicators about whether the mother and the father completed at least primary education, whether the respondent is married, and a household asset index. Detailed definitions of these variables are provided in Appendix A.4.

The parameter of interest is β, the average difference between treatment and control observations at end-line and/or follow-up. Under the assumption that the control observations constitute a valid counterfactual for the treatment sample, this measure is the intent to treat (ITT) estimate, which identifies the causal effect of the programme on parents and child that attended the parenting sessions.

Since the randomization was implemented at the sector level and observations might be correlated within clusters, we clustered the standard errors at this level. However, as the number of clusters is small (nine clusters), our standard errors might be biased downwards (Bertrand et al. 2004). Therefore, our statistical inference is based on a bootstrap t-test using the wild-cluster bootstrap procedure which allows us to estimate precise estimations of p-values with less than ten clusters, as in our case (Cameron et al. 2008; Roodman et al. 2019).

The presence of multiple outcomes in this study creates the potential problem of cherry-picking significant estimates, and the need to correct for multiple hypothesis testing. We use the Romano and Wolf (2005) correction to address the possibility of arbitrarily selecting statistically significant treatment effects, and to reduce the likelihood of family wise type-I error (FWER). Appendix C.4 provides a detailed description of this correction.

5 Findings

In this section, we present the results of the evaluation of the impact of First Steps on child development and parental outcomes at end-line (12 months after baseline) and follow-up (33 months after baseline).

5.1 Short-term impact (12 months)

Child development

Table 1 reports the estimated coefficients of the impact of the First Steps programme on child development outcomes 12 months after baseline. Both LT and FT interventions show a positive impact, with the effect size ranging between 0.25 and 0.48 standard deviations. Column 6 shows that the programme increased the aggregate index of child development by 0.29 and 0.38 standard deviations, respectively, for LT and FT interventions. The FT intervention arm has a stronger effect in all five child development dimensions, with the exception of the gross motor skills outcome, where the effect for FT is not significant. A similar result was found in other early child development programme evaluations (see for example Martinez et al. 2017), possibly indicating that gross motor skills may take a longer time to develop in children in these age groups. The overall effects are in line with other studies that use ASQ to assess child development outcomes, which report effects ranging between 0.20 and 0.40 SD (Doyle et al. 2017; Martinez et al. 2017).

The largest effects are concentrated in communication, problem solving, and personal social skills. This is not surprising, since the First Steps curriculum strongly highlighted these areas, with a particular emphasis, given the context in which it was implemented, on early communication skills, such as talking, singing, playing, reading, story-telling, bonding, and touching. This finding is also consistent with parents reporting mostly their participation in sessions focusing on these topics (see Table B2).[22]

Parental time investment

The results in Table 2 show a positive and significant effect of the programme on maternal time investment in the short term, with the effect mostly driven by learning caregiver-child activities. The effect sizes of the average aggregate index are large, ranging between 0.4 SD (LT group) and 0.6 SD (FT group). Table B5 shows also a positive impact of the programme on fathers' time investment, in particular on learning and positive discipline activities. It is to note that the FT group also experienced a significant impact on learning father-child activities. This result is consistent with the results in column 1 of Table B5, and with evidence from a qualitative study on First Steps, which highlighted the role of the home visits (implemented only in the FT group) in promoting the engagement of fathers with their children (Fuller 2016). As discussed before, however, observations about fathers are small and, hence, results for paternal time investment need to be taken with caution.[23]

[22] In addition, Table B4 shows that child development outcomes, mother time investment and mothers' influence are positively correlated with the number of sessions attended, with the effect fading out in the medium term for child development and with a smaller but still positive and significant estimate for the maternal time investment and maternal influence.

[23] We analysed the correlation between who in the family participated in First Steps and the outcomes of interest. We defined a categorical variable equal to 0 if the caregiver did not participate (baseline group: control group and non-compliers), equal to 1 if the father participated alone, equal to 2 if the mother participated alone, and equal to 3 if they participated together. The effect of the programme is larger if both the mother and father attended First Steps (Panel A Table B6). It is to note, however, that the effect fades out at follow-up, with the exception of maternal time investment (Panel B Table B6).

We also present estimates related to each individual activity performed by mothers.[24] These results are reported in Table B7. The programme's largest impact on maternal time investment is related to activities such as singing, telling a story, playing with toys, reading, and counting. This is consistent with the focus of the key contents of the parent group-meeting sessions.[25] As we discussed in the introduction, the research design used to evaluate First Steps does not allow us to disentangle the individual effects of the various components of the intervention. This result suggests, nonetheless, that the book gift component may have had an important effect. Future research on First Steps or other early child development interventions in other low-income contexts should investigate further the potential of book gifts as a form of alleviating education and financial barriers to how low-income mothers and fathers engage in their children's development from an early age. Other early child development programmes where sessions were built on a similar *curriculum* have found positive results from similar activities (Engle et al. 2011; Knauer et al. 2016).

Parental influence

We investigated the impact of First Steps on perceived parental influence over children's (i) learning, (ii) development, (iii) nutrition, (iv) care, (v) discipline or child guidance, and (vi) health care. Results in Table 3 show that both the LT and FT intervention arms led to increases in how mothers perceived their influence over their children's development by 0.4 SD and 0.6 SD, respectively. These are large effects and, similarly to the results above, are more pronounced for families in the FT arm. We observe similar effects of First Steps on how fathers perceive their influence over the children's development (Table B8).

These results suggest that First Steps has led to substantial changes in parental behaviour and beliefs, which in turn may have reinforced the impact of the programme on child development outcomes. We return to this issue later in the paper.

5.2 Medium-term impact (33 months)

Child development

Table 4 shows the estimated coefficients of the programme's impact after almost three years from baseline (33 months). The results show that the effect of the FT intervention on child development outcomes, despite smaller (between 0.13 and 0.24 SD) than the results discussed above and statistically weaker, persist in the longer term. The LT intervention does not show a significant effect over the longer term.

The evidence on the longer-term effect of early child development programmes is mixed with some studies reporting a fading effect (for a review, see Bailey et al. 2017), whilst other programme evaluations find ITT effects ranging from 0.1SD to 0.2SD of a standard deviation after three years on similar child development outcomes (Carneiro et al. 2019a). Also in line with our results are those programmes that used the same tool (the ASQ) to evaluate child development outcomes. These have found results ranging between 0.2 SD and 0.35 SD after two or three years (Doyle et al. 2017; Martinez et al. 2017).

---

[24] We have obtained similar estimates for fathers. They are available upon request. In the paper, we show only the main tables for father-related outcomes given the low number of observations.

[25] The read and show books activities have a strikingly large effect. This might be due to the fact that, as part of the First Step programme in the FT group, one book was gifted to parents. The aggregate mean index we constructed does not include the read and show book activities. Results remain consistent if we include these activities, but the magnitudes of the effects becomes larger. Results are available upon request.

This result is informative from a policy point of view. It suggests that nearly three years after the programme's implementation, children of highly economically vulnerable parents that participated in First Steps show improvements in their development outcomes, largely in their communication, gross motor, and personal social skills. These results suggest that the additional components in the FT arm—the extra facilitator, the home visit, the leaflets, and the book—made a positive difference in producing sustainable impacts over time in these remote rural communities. Due to the design of the programme's implementation, we are not able to disentangle and estimate the causal impact of each of these additional components of the FT intervention. However, this result is in line with studies reporting advantages of home visits and book provision elsewhere in the (more developed) world (Grantham-McGregor et al. 1991; Walker et al. 2005; Gertler et al. 2014).

Parental time investment

Table 5 shows the effects of the programme on maternal time investment outcomes in the longer term. Although the effect sizes are smaller than in the 12-month evaluation, we find that the impact of First Steps on maternal time investment is robust and persists 33 months later. Both the LT and FT groups observe a positive and significant impact of the programme on maternal time investment (0.2 SD). As in the previous section, paternal time investment results (Table B9), show that only the FT effect persists in the longer term. The activities that show the largest impact are the learning activities, a result that mirrors the shorter-term results analysed above.

Table B10 shows the effects of the programme in the longer term on each activity conducted by the mother. Playing, singing, counting, reading, and teaching something new still show the largest impacts for mothers, in line with the shorter-term estimates. This is also in line with previous studies. For example, Knauer et al. (2016), in a group-based programme in Mexico, finds that after the programme—in the medium term—parents were more likely to engage in playing, storybook reading, and singing. Similar results are discussed in the systematic review conducted by Engle et al. (2011) .

Other parental outcomes

We investigate here the impact of First Steps on parental self-efficacy (PSE). Some recent studies have looked at parental self-efficacy as an additional outcome of interest (Doyle et al. 2017; Attanasio et al. 2018; Carneiro et al. 2019a; Chandra et al. 2020). PSE has been studied as an indicator of successful treatment (Tucker et al. 1998; Hoza et al. 2000) and as a mechanism for parenting behavioural changes targeted by early child development interventions (Spoth et al. 1995; Miller-Heyl et al. 1998).

Results in Table 6 show that the FT intervention arm has resulted in increases in mother self-efficacy aggregate mean index by 0.14 SD (column 9).[26] The table shows that First Steps, and in particular the FT intervention, affected play, control, discipline, pressure, and learning dimensions of self-efficacy, with the exception of emotion, empathy, and self-acceptance. These results corroborate the findings of similar early child development interventions on the PSE sub-scales, analysed with the TOPSE questionnaire (Bloomfield and Kendall 2012; Ulfsdotter et al. 2014; Enebrink et al. 2015; Miller and Harrison 2015) and on parental self-efficacy in general, measured using different tools (Carneiro et al. 2019a). An early study by Mondell and Tyler (1981) found that parents with higher levels of self-efficacy provided more help, gave fewer commands, and showed more positive influence as they interacted with their children than those with lower levels of self-efficacy.

Parents who have strong efficacy beliefs are also likely to promote educational activities conducive to learning (Bandura et al. 1996). They are also better able to influence their children's skills development

[26] Wwe do not have information at baseline on these outcomes, as we collected this data only at follow-up, and therefore cannot control for baseline values in the estimated model. However, the relevance of these results merits its own analysis.

than those parents who doubted their ability to influence their children's development (Schneewind and Pfeiffer 1995; Barlow et al. 2012; Dowling 2014).

As mentioned in Section 3, we included a number of questions on parental attitudes at follow-up (see Appendix A.3 for a description of the variables). We focused on parental attitudes across four dimensions: child health, child behaviour, gender roles in the household, and child development. Interestingly, results in Table 7 show that First Steps has a positive effect on mother attitudes over gender roles in the household, for both LT and FT groups (0.11 SD and 0.17 SD, respectively).

We further disentangled the effects of the programme on each component of attitudes towards gender roles. We found that the strongest impact is related to gender attitudes towards child care and household chores (see table B11). Other attitudinal dimensions are not statistically significant, and the maternal attitude aggregate index is not significant. Taken together, these results suggest that the intervention not only improved child development outcomes and parenting engagement, but has also led to changes in normative attitudes that may lead in the future to reductions in gender gaps in child development. This could be an important topic of analysis in future research and policy design of early child development interventions.

We also asked questions about locus of control. Individual behaviours can be substantially driven by believes. If a person is not convinced that their actions will make a difference, the resources and opportunities that may be available will remain unexploited. In the psychological literature, these beliefs are understood as 'locus of control'. Having a high locus of control has been shown to be positively related to job performance, schooling decisions, employment, and occupational choice (Judge and Hurst 2007). We focus on the internal locus of control in particular—the beliefs that one's life outcomes are controlled by oneself—as opposed to the external locus of control—life outcomes depend on exogenous forces such as fate or luck. Results in Table 8 show that First Steps had a positive effect on mother's internal locus of control for both LT and FT groups (0.25 SD and 0.23 SD, respectively).

Lastly, we analysed the aspiration gap of the caregiver and then the caregiver's aspiration towards their child. The presence of forward-looking goals (i.e. aspirations) is an important concept taken from psychology. A weak capacity to aspire can translate into low or no investments. Aspirations and individual behaviour are connected only when the difference between the conditions in which the person lives and the standard of living the person aspires to have is taken into account (Ray 2006). This aspiration gap (rather than the aspiration on its own) is the driver of future behaviours.[27] Results in Table 8 show that the FT treatment affected the aspiration gap of the mother: whether she has a role model and can become like her role model within five years (0.18 SD), and whether she would like to change jobs in order to achieve a better life (0.19 SD). However, the effects of the FT and LT treatments are not statistically different from each other. The FT had a significant and positive impact on aspirations of mothers that their child will obtain a university degree. Both LT and FT show a positive but statistically insignificant result on the child's preferred age of marriage.

Taken together, results on the effect of First Steps on parental attitudes, locus of control, and aspirations suggest a sustained effect of First Steps, not only on child development, parental time investments, and confidence in rural, remote, and poor communities, but also on other fundamental attitudes and beliefs that may sustain the effects of the programme over the longer term. Arguably, it may have important spillover effects on other household development outcomes.

[27] See Dercon et al. (2012) for a discussion on beliefs, locus of control, and aspirations.

6 Heterogeneous treatment effects

Existing studies have reported differences in early child development outcomes resulting from parenting interventions in terms of, for instance, gender (Heckman et al. 2010; Doyle et al. 2017; Doyle 2020), age (Heckman et al. 2010; Conti et al. 2016), and household socioeconomic characteristics (Doyle 2020). In this section, we analyse similar heterogeneous effects by testing whether the impact of First Steps differs across the gender and age of the child, assets held by the family, the level of education of the mother and father, and baseline levels in the child development aggregate index. We report results on heterogeneous effects estimated 12 months (Table 9) and 33 months after baseline (Table B12). In each table, we report heterogeneity effects on the child development aggregate index in Panel A. In Panel B and C, we show the heterogeneous impacts on maternal time investment and maternal influence aggregate indices, respectively. Each row reports estimates from an OLS regression. In columns 1 and 2 we report the uninteracted effect of LT and FT treatments, respectively. Columns 3 and 4 report the interaction between LT (FT) treatments and the characteristics of interest.

Gender. Results in Panel A of Table 9 show that the programme's effect on child development is the same regardless of the gender of the child. This result mirrors other studies (Doyle 2020), although sources of gender differentiated responses to early child development programmes is not yet well understood (Conti et al. 2016), and may vary depending on the specific stage of the life cycle and the particular measure used (Matthews et al. 2009; Doyle 2020). In the longer term, results in Panel A of Table B12 show that the effect of the programme is stronger if the child is a girl, similar to other studies (Sandner and Jungmann 2017). This may well have to do with changes observed in caregiver attitudes towards gender norms, as discussed above, which may have led to more attention being paid to girls. However, this is only speculative, and more research is needed to better understand the gender effects of group-based parenting interventions, such as First Steps, and their transmission over time. Panel B and C of Table 9 and Table B12 do not show any differentiated effects by gender in parental outcomes.

Age of child. Results in Table 9 do not show differences in the effect of the programme on child development and parental outcomes as the child age increases. Other studies find that the effect of early child development interventions on child development is larger as the child gets older (Banerji et al. 2017). However, as with gender differences, these results are not well understood, and may depend on particular community and household characteristics that may favor certain age groups over others, as well as the share of younger and older siblings in the household. We find also that, at follow-up, the effects on child and parental outcomes do not differ by child's age (see Table B12).

Household assets. We analysed whether the effect of the programme differs across levels of household wealth. We interacted the treatment variables with an asset index indicator equal to 1 if a household's asset index is above the median level, and 0 if it is below. The results do not show any differentiated effect of the programme by household's wealth status on any of the outcomes of interest. The existing evidence on the heterogeneous impacts of early child development interventions by wealth status is mixed. Doyle (2020) finds that a programme introduced in Ireland benefited mostly poorer families. In contrast, Carneiro et al. (2019a) and (Blimpo et al. 2019), who study the impact of early child development programmes targeted at poor families in Chile and The Gambia, respectively, do not find any differences in the impact of the programmes along wealth status. These results, like ours, are probably explained by the fact that wealth differences between poor families and households are not large. Thus, small differences in wealth within the same locality may, therefore, not influence the impact of early child development interventions much.

Caregiver education. We tested whether the impact of the programme varies with the level of education of the mother or the father. The results show that the impact of the programme on child development outcomes and on maternal time investment is similar across families with different levels of maternal

and paternal education. However, the impact of the programme on maternal influence is smaller in families where the mother or the father are more educated. This result is consistent with other findings in the literature where parents education in the short term is either not a significant factor (Banerji et al. 2017; Baranov et al. 2020), or the treatment effects is larger among less educated parents (Carneiro et al. 2019a). The effects fade away in the follow-up survey (see Table B12), suggesting a convergence of outcomes across household education levels.

Child development at baseline. Finally, we analysed whether the impact of the programme changes across different levels in child development at baseline. To do this, we separated the child development aggregate index at baseline in quartiles, and interacted each quartile with each of the two treatment binary variables (LT and FT). We defined the upper quartile as the reference category. Results in Table 10 show that the impact of First Steps on child development, maternal time investment, and maternal influence is smaller for children who were in the upper quartile in the child development index at baseline. However, the effect on child development is significant only for the LT group in the highest quartile of child development. Both LT and FT arms have a smaller impact on maternal time investment in children that are better off in their child development index at baseline. This result is consistent with recent findings in the literature (Warrinnier et al. 2018).

Over time, this smaller impact among children that were better off at baseline disappears, suggesting a convergence in child development outcomes over time (Table B13). Similar result was found in Carneiro et al. (2019a).

7 Mechanisms and mediation analysis

Recent studies have shown the importance of the role of parents and their behaviour as mediating the impact of parenting interventions on child development outcomes. They find that changes in child development outcomes are largely explained by changes in parental investments and strengthening of positive parenting practices (Carneiro et al. 2019a; Attanasio et al. 2020).

In this section, we discuss two potential parenting mechanisms through which the First Steps programme may have affected child development outcomes: changes in parental time investment and changes in how parents perceive their ability to influence the development of their children.[28]

The direction of the mediating effect of parental time investments is not clear a priori. The First Steps programme aimed to strengthen child-parent interactions, and to encourage parents to invest more quality time with their children. However, such effects may not materialize if the time the parents spend with their First Steps child decreases because they perceive the intervention itself as some form of time investment in that child, and they shift their attention and resources to other children in the household.[29] This type of crowding out is discussed in Attanasio et al. (2020). We tested for a potential crowding out effect by including in our specification an interaction between a variable that reports the number of children below three in the household and the treatment dummies. Table B14 shows the coefficients on the interaction terms are not significant at end-line (Panel A), and small and weakly significant at follow-up, both for child development and mother self-efficacy outcomes (Panel B). These results are suggestive that crowding out may not be a key mechanism at play in our study.

[28] We acknowledge that other parental mechanisms may also mediate the changes observed on child development. In this study, we focus on parental mechanisms that we explicitly measured in the survey.

[29] Note that the time investments of parents into the First Steps programme was of moderate duration (two hours, once a week for 17 weeks).

Another possible channel through which First Steps could influence child development outcomes is through an increase in parental confidence in their influence over their children's development trajectories. The effect of this potential mediator factor is again undetermined a priori. First Steps throughout the radio episodes and other activities constantly reminded parents about their importance in the development of the child, hence possibly boosting their confidence and influence. First Steps weekly meetings were also implemented at the village level in groups. Group settings may improve parental confidence by creating peer support effects and positive changes in group social norms (Doyle 2020). It is nonetheless possible that group settings may reduce parents' confidence (and hence how they perceive their influence over the child's development) if they feel less competent than other parents attending the group meetings (Andrew et al. 2018). Table B15 shows the correlation of the mediator factors with child development indicators. It indicates that maternal time investment is highly correlated with child development outcomes. It also shows that maternal influence is correlated too, but with lower economic and statistical relevance.[30]

In order to shed further light on the potential mechanisms at play, we decomposed the effects of the intervention on child development outcomes into components attributable to changes in maternal time investments and mother's perceived influence on their children, using a liner mediation analysis (Cunha and Heckman 2007; Heckman et al. 2013; Kim et al. 2018; Carneiro et al. 2019a).

Findings in earlier sections of the paper show that First Steps resulted in large positive impacts on parental time investment, parents' perceived influence over their child, and child development outcomes. The effect of the programme on child development outcomes is mediated through changes in parental behaviour only if experimentally induced changes to parents' behaviour correlate with experimentally induced changes in a child's outcome (Kim et al. 2018). In order to establish which of these observed parental inputs may explain the observed impacts on child outcomes, we decomposed the intervention effect of child development into changes in maternal time investment, mother's perceived influence, and other unobserved factors. These unobserved factors include any change that is not captured by our indicators of changes in parenting outcomes. These could include the experimentally induced improvement in parental confidence, in parental locus of control, in parental self-efficacy, in parental mental health, and in material investment (such as buying toys or books), among others. We conducted a linear mediation analysis, and estimated a linear equation model to decompose the treatment effect. The equation and the estimation process are described in Appendix C.6. The decomposition of the treatment effects shows that experimentally induced improvements in parenting explains the impact of the intervention on parental outcomes (i.e. maternal time investment and perceived parental influence over their child's development).

Figure 3 shows the estimates of the mediation analysis at end-line (12 months). The results show that the increase in maternal time investment explains almost 20 per cent of the impact of the intervention on the child development index. Decomposing the child development index in its components, we observe that maternal time investment contributes to around (i) 20 per cent increase in communication and problem solving skills; (ii) 60 per cent increase in gross-motor skills; (iii) five per cent increase in fine-motor skills; (iv) and about 30 per cent of the increase in personal-social skills. Changes in mother's perceived influence explain a much smaller proportion of the overall effect at 12 months—by around five per cent of the effect in all the dimensions and in the child development mean index. The only exception is the fine-motor skills, where 15 per cent of the increase is explained by maternal influence, in contrast to only five per cent increase explained by maternal time investment.

[30] We excluded paternal time investment from this correlational analysis to avoid reducing the sample to 87 observations. We have however estimated the same regressions using paternal time investment as reported by the mother, and results, available upon request, suggest that maternal time investment explains most of the correlation.

We also performed the mediation analysis for the follow-up (33 months) results. Figure B1 shows that results at 33 months are consistent with those at 12 months. The main difference is that the increase in maternal time investment explains more than 35 per cent of the treatment effect on child development outcomes. The largest contribution of the increase in maternal time investment is on problem solving skills.[31] These results, taken together with the 12-month analysis, suggest the importance of maternal time investments in ensuring the positive impact of First Steps on child development outcomes and their sustainability across time.

8 Conclusions

This paper analysed the impact of First Steps, a group-based early child development programme that attempted to improve parenting skills among families of children aged six to 24 months in a remote and poor district in Rwanda. These families constitute some of the most economically vulnerable people in the world. Much of the evidence available on the effect of early child development interventions results from programmes implemented in the USA, with some emerging evidence on disadvantaged populations in medium-income countries. To date, few detailed studies occurred in low-income rural contexts, where families in remote and isolated settings face serious financial, social, emotional, and physical constraints that may crowd out any attempts to improve parenting skills and early child development outcomes. This paper provides one of the first studies in such a setting. We investigated the impact of First Steps 12 months and 33 months after its implementation, and examined its effects on five dimensions of child development, on mother and father's time investments, and on the perceived influence of mothers and fathers over their child development pathways. During the follow-up survey, we investigated additional parental outcomes that may influence parental behaviours and, in turn, child development outcomes, including parental self-efficacy, parental attitudes towards child development and gender norms, locus of control, and parental aspirations about themselves and about their children.

The First Steps programme was designed as a cluster-randomized controlled trial with three intervention arms: a control group, a light treatment group, and a full treatment group. The results show that, twelve months after baseline, children in families that participated in First Steps benefited from substantial improvements in their development outcomes, in particular in communication, problem solving, and personal social skills. Children in families in the full treatment group showed the largest improvements in the child development aggregate index (0.4 SD). These positive effects persisted 33 months later (0.2 SD).

In addition, the results show that First Steps had a strong and positive impact on parental outcomes. The programme increased the reported frequency of activities in which mothers and fathers engage with their child, mostly through play and learning mother-child activities. The programme also positively affected how parents perceive their ability to influence their children's development outcomes. These effects are stronger for families in the full treatment group, and persist in the longer term, albeit with smaller effects.

Finally, we find positive changes on parental self-efficacy, locus of control, attitudes towards gender norms, and aspirations. We investigated what potential mediating factors may explain these results, and found that, after 12-months, around 20 per cent of the positive changes in child development can be attributed to the increase in the reported frequency of activities mothers engage with their children. A smaller proportion of changes in child development outcomes can be attributed to programme-induced changes in how parents perceive their ability to influence their children's development, which may

[31] These results for the medium term need to be taken with caution as the test to validate this analysis failed (more details in Appendix C.6).

have improved parents' confidence and awareness of their importance in their child's lifelong development.

Taken together, these results show that First Steps induced positive changes in child development outcomes and in parental behaviour, attitudes, and beliefs among very poor families in rural Rwanda, 12 months after the implementation of the programme. They were still observed (albeit with a lower coefficient) almost three years after the implementation of the programme. These effects were more pronounced in the full treatment group, which, in addition to listening to a radio programme and attending group discussions, also benefited from an extra trained facilitator, one home visit, leaflets to reinforce the message of the group discussions, and the gift of a book. We are not able to disentangle the effect of individual components of the programme, but suspect these results may well be due to the emphasis placed on small barrier-breaking nudges, such as ideas on how parents can spend quality time with children while conducting daily chores, the use of local and easily available materials, a home visit, and the provision of books. These findings generate important insights for future early child development programmes in poor rural areas, as well as future research questions in terms of refining how human capital production functions may operate in these contexts. Furthermore, there is a need to continuously follow the children in First Steps and their parents across time. The programme may yield further valuable information about how permanent observed changes in attitudes and behaviour may have continued across time, about their effects as children join school and become teenagers and adults, and about possible breaks in poverty traps if First Steps children are enabled to get better education and jobs than their parents (see Heckman et al. 2010; Campbell et al. 2014).

This study was based on a localized intervention in a specific area of Rwanda. Thus, results may not be generalized to other settings. However, the low intensity of the programme, combined with its impressive sustained effects over time despite the serious constraints faced by the families targeted, suggests that there may be substantial benefits in scaling up the intervention across Rwanda and in other low-income settings, where more time consuming and intense interventions may not be viable. Further interventions and research in such settings is urgently needed as child poverty is increasingly being concentrated in a handful of very low-income countries, such as Rwanda (Shepherd et al. 2014). If the Ngororero results generalize to other similar settings, interventions such as First Steps may represent a valuable means to break the persistence of poverty traps across generations of parents and children deprived of adequate parenting knowledge and confidence.

References

Abimpaye, M., Dusabe, C., Nzabonimpa, J. P., Ashford, R., and Pisani, L. (2019). 'Improving parenting practices and development for young children in rwanda: Results from a randomized control trial'. *International Journal of Behavioral Development*, *Prepublished*(OnlineFirst): 16 July 2019. doi: 10.1177/0165025419861173

Aboud, F. E., and Yousafzai, A. K. (2015). 'Global Health and Development in Early Childhood'. *Annual Review of Psychology*, *66*(1): 433–57.

Almond, D., Currie, J., and Duque, V. (2018). 'Childhood Circumstances and Adult Outcomes: Act II'. *Journal of Economic Literature*, *56*(4): 1360–446.

Anderson, M. L. (2008). 'Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects'. *Journal of the American statistical Association*, *103*(484): 1481–95.

Andrew, A., Attanasio, O., Bernal, R., Sosa, L. C., Krutikova, S., and Rubio-Codina, M. (2019). 'Preschool quality and child development'. NBER Working Paper 26191. Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w26191

Andrew, A., Attanasio, O., Fitzsimons, E., Grantham-McGregor, S., Meghir, C., and Rubio-Codina, M.

(2018). 'Impacts 2 years after a scalable early childhood development intervention to increase psychosocial stimulation in the home: A follow-up of a cluster randomised controlled trial in Colombia'. *PLOS Medicine*, *15*(4): 1– 19. doi: 10.1371/journal.pmed.1002556

Attanasio, O., Baker-Henningham, H., Bernal, R., Meghir, C., Pineda, D., and Rubio-Codina, M. (2018). 'Early stimulation and nutrition: The impacts of a scalable intervention'. NBER Working Paper 25059. Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w25059

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

Attanasio, O., Meghir, C., and Nix, E. (2015). 'Human capital development and parental investment in india'. NBER Working Paper 21740. Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w21740

Bailey, D., Duncan, G. J., Odgers, C. L., and Yu, W. (2017). 'Persistence and fadeout in the impacts of child and adolescent interventions'. *Journal of Research on Educational Effectiveness*, *10*(1): 7–39.

Bandura, A., Barbaranelli, C., Caprara, G. V., and Pastorelli, C. (1996). 'Multifaceted Impact of Self-Efficacy Beliefs on Academic Functioning'. *Child Development*, *67*(3): 1206–22.

Banerjee, A., Ferrara, E. L., and Orozco-Olvera, V. H. (2019). 'The entertaining way to behavioral change: Fighting hiv with mtv'. NBER Working Paper 26096. Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w26096

Banerji, R., Berry, J., and Shotland, M. (2017). 'The Impact of Maternal Literacy and Participation Programs: Evidence from a Randomized Evaluation in India'. *American Economic Journal: Applied Economics*, *9*(4): 303–37.

Baranov, V., Bhalotra, S., Biroli, P., and Maselko, J. (2020). 'Maternal Depression, Women's Empowerment, and Parental Investment: Evidence from a Large Randomized Control Trial'. *American Economic Review*, *110*(3): 824–59.

Barlow, J., Smailagic, N., Huband, N., Roloff, V., and Bennett, C. (2012). 'Group-based parent training programmes for improving parental psychosocial health'. *Campbell Systematic Reviews*, *8*(1): 1–197.

Bayley, N. (2006). *Bayley scales of infant and toddler development* (3rd ed.). San Antonio, TX: Pearson.

Bernard, T., Dercon, S., Orkin, K., and Seyoum Taffesse, A. (2015). 'Will video kill the radio star? Assessing the potential of targeted exposure to role models through video'. *World Bank Economic Review*, *29*(suppl_1): 226–37.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). 'How Much Should We Trust Differences-in-Differences Estimates?' *Quarterly Journal of Economics*, *119*(1): 249–75.

Black, M. M., Walker, S. P., Fernald, L. C., Andersen, C. T., DiGirolamo, A. M., Lu, C., McCoy, D. C., Fink, G., Shawar, Y. R., Shiffman, J., et al. (2017). 'Early Childhood Development Coming of Age: Science Through the Life Course'. *Lancet*, *389*(10064): 77–90.

Blimpo, M. P., Carneiro, P. M., Jervis, P., and Pugatch, T. (2019). 'Improving Access and Quality in Early Childhood Development Programs: Experimental Evidence from The Gambia'. GLO Discussion Paper 318. Essen: Global Labor Organization.

Bloomfield, L., and Kendall, S. (2012). 'Parenting self-efficacy, parenting stress and child behaviour before and after a parenting programme'. *Primary Health Care Research and Development*, *13*(4): 364–72.

Bradley, R. H., and Caldwell, B. M. (1977). 'Home Observation for Measurement of the Environment: a Validation Study of Screening Efficiency.' *American Journal of Mental Deficiency*, *81*(5): 417–20.

Britto, P. R., Lye, S. J., Proulx, K., Yousafzai, A. K., Matthews, S. G., Vaivada, T., Perez-Escamilla, R., Rao, N., Ip, P., Fernald, L. C., et al. (2017). 'Nurturing Care: Promoting Early Childhood Development'. *Lancet*, *389*(10064): 91–102.

Britto, P. R., Ponguta, L. A., Reyes, C., and Karnati, R. (2015). *A Systematic Review of Parenting Programmes for Young Children*. New York: UNICEF.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). 'Bootstrap-based Improvements for Inference with Clustered Errors'. *Review of Economics and Statistics*, *90*(3): 414–27.

Campbell, F. A., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E., and Pan, Y. (2014). 'Early Childhood Investments Substantially Boost Adult Health'. *Science*, *343*(6178): 1478–85.

Campbell, F. A., and Ramey, C. T. (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.

Carneiro, P. M., Galasso, E., Lopez Garcia, I. X., Bedregal, P., and Cordero, M. (2019a). 'Parental Beliefs, Investments, and Child Development: Evidence from a Large-Scale Experiment'. IZA Discussion Paper 12506. Bonn: Institute of Labor Economics.

Carneiro, P. M., Kraftman, L., Mason, G., Moore, L., and Rasul, M., I. and. Scott. (2019b). 'The Impacts of a Multifaceted Pre-natal Intervention on Human Capital Accumulation in Early Life'. Working Paper. London: University College London.

Cattaneo, M. D. (2010). 'Efficient semiparametric estimation of multi-valued treatment effects under ignorability'. *Journal of Econometrics*, *155*(2): 138–54.

Chandra, A., Mani, S., Dolphin, H., and Dyson, M. (2020). *Impact Evaluation of an Integrated Early Childhood Parenting Program in Sierra Leone.* (Unpublished)

Chetty, R., Hendren, N., and Katz, L. F. (2016). 'The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment'. *American Economic Review*, *106*(4): 855–902.

Clarke, D., Romano, J. P., and Wolf, M. (2019). 'The Romano-Wolf Multiple Hypothesis Correction in Stata'. IZA Discussion Paper 12845. Bonn: Institute of Labor Economics.

Conti, G., Heckman, J. J., and Pinto, R. (2016). 'The effects of two influential early childhood interventions on health and healthy behaviour'. *Economic Journal*, *126*(596): F28–F65.

Cunha, F., Elo, I., and Culhane, J. (2013). 'Eliciting maternal expectations about the technology of cognitive skill formation'. NBER Working Paper 19144. Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w19144

Cunha, F., and Heckman, J. J. (2007). 'The Technology of Skill Formation'. *American Economic Review*, *97*(2): 31–47.

Currie, J., and Almond, D. (2011). 'Chapter 15 - Human Capital Development Before Age Five'. In D. Card and O. Ashenfelter (eds), *Handbook of labor economics* (Vol. 4 Part B, pp. 1315–486). Amsterdam: Elsevier. doi: 10.1016/S0169-7218(11)02413-0

Delavande, A., Giné, X., and McKenzie, D. (2011). 'Measuring subjective expectations in developing countries: A critical review and new evidence'. *Journal of Development Economics*, *94*(2): 151–63.

Dercon, S., Bernard, T., Taffesse, A. S., et al. (2012). 'Beyond Fatalism an Empirical Exploration of Self-efficacy and Aspirations Failure in Ethiopia'. ESSP working papers 46. Washington D.C.: International Food Policy Research Institute.

Doepke, M., Sorrenti, G., and Zilibotti, F. (2019). 'The Economics of Parenting'. *Annual Review of Economics*, *11*(1): 55–84. doi: 10.1146/annurev-economics-080218-030156

Doepke, M., and Zilibotti, F. (2017). 'Parenting with style: Altruism and paternalism in intergenerational preference transmission'. *Econometrica*, *85*(5): 1331–71.

Dooley, M., and Stewart, J. (2007). 'Family income, parenting styles and child behavioural–emotional outcomes'. *Health Economics*, *16*(2): 145–62.

Dowling, H. (2014). *Parental self-efficacy in early years parenting* (Unpublished doctoral dissertation). University of Manchester.

Doyle, O. (2020). 'The First 2,000 Days and Child Skills'. *Journal of Political Economy*, *Prepublished*(online): 16 April 2020. doi: 10.1086/705707

Doyle, O., Harmon, C., Heckman, J. J., Logue, C., and Moon, S. H. (2017). 'Early skill formation and

the efficiency of parental investment: A randomized controlled trial of home visiting'. *Labour Economics*, *45*(-): 40–58.

Enebrink, P., Danneman, M., Benvestito Mattsson, V., Ulfsdotter, M., Jalling, C., and Lindberg, L. (2015). 'ABC for Parents: Pilot Study of a Universal 4-Session Program Shows Increased Parenting Skills, Self-efficacy and Child Well-Being'. *Journal of Child and Family Studies*, *24*(7): 1917–31. doi: 10.1007/s10826-014-9992-6

Engle, P. L., Fernald, L. C., Alderman, H., Behrman, J., O'Gara, C., Yousafzai, A., de Mello, M. C., Hidrobo, M., Ulkuer, N., Ertem, I., et al. (2011). 'Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries'. *Lancet*, *378*(9799): 1339–53.

Fernald, L. C., Galasso, E., Qamruddin, J., Ranaivoson, C., Ratsifandrihamanana, L., Stewart, C. P., and Weber, A. M. (2019). 'Effects of nutritional supplementation and home visiting on growth and development in young children in Madagascar: a cluster-randomised controlled trial'. *Lancet: Global Health*, *16*(1): 466.

Fuller, R. Y. (2016). 'Qualitative evaluation of Save the Children's First Steps program in Rwanda'. Working Paper. University of Wisconsin.

Gertler, P., Heckman, J. J., Pinto, R., Zanolini, A., Vermeersch, C., Walker, S., Chang, S. M., and Grantham-McGregor, S. (2014). 'Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica'. *Science*, *344*(6187): 998–1001.

Glennerster, R., and Takavarasha, K. (2013). *Running Randomized Evaluations: A Practical Guide*. Princeton, NJ: Princeton University Press. doi: 10.2307/j.ctt4cgd52

Gollenberg, A. L., Lynch, C., Jackson, L., McGuinness, B., and Msall, M. (2010). 'Concurrent validity of the parent-completed Ages and Stages Questionnaires, with the Bayley Scales of Infant Development II in a low-risk sample'. *Child Care, Health and Development*, *36*(4): 485–90.

Grantham-McGregor, S. M., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., Strupp, B., Group, I. C. D. S., et al. (2007). 'Developmental Potential in the First 5 years for Children in Developing Countries'. *Lancet*, *369*(9555): 60–70.

Grantham-McGregor, S. M., Powell, C. A., Walker, S. P., and Himes, J. H. (1991). 'Nutritional Supplementation, Psychosocial Stimulation, and Mental Development of Stunted Children: the Jamaican Study'. *Lancet*, *338*(8758): 1–5.

Heckman, J. J. (2006). 'Skill Formation and the Economics of Investing in Disadvantaged children'. *Science*, *312*(5782): 1900–02.

Heckman, J. J., and Carneiro, P. M. (2003). 'Human Capital Policy'. NBER Working Paper 9495. Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w9495

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P., and Yavitz, A. (2010). 'Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program'. *Quantitative Economics*, *1*(1): 1–46.

Heckman, J. J., Pinto, R., and Savelyev, P. (2013). 'Understanding the mechanisms through which an influential early childhood program boosted adult outcomes'. *American Economic Review*, *103*(6): 2052–86.

Hoza, B., Owens, J. S., Pelham, W. E., Swanson, J. M., Conners, C. K., Hinshaw, S. P., Arnold, L. E., and Kraemer, H. C. (2000). 'Parent cognitions as predictors of child treatment response in attention-deficit/hyperactivity disorder'. *Journal of Abnormal Child Psychology*, *28*(6): 569–83.

IMF. (2019). *World Economic Outlook.* Washington, D.C.: International Monetary Fund.

Judge, T. A., and Hurst, C. (2007). 'Capitalizing on one's advantages: Role of core self-evaluations.' *Journal of Applied Psychology*, *92*(5): 1212.

Justino, P., Leone, M., Rolla, P., Abimpaye, M., Germond, R., Malik, S., and Uwamahoro, D. (unpublished). 'Nudging Parental Investment by Improving Efficacy and Beliefs? A Randomized Video Intervention in Rwanda'. Working Paper. (Institute of Development Studies and University of Pavia)

Katz, L. F., Kling, J. R., and Liebman, J. B. (2001). 'Moving to opportunity in Boston: Early results of

a randomized mobility experiment'. *Quarterly Journal of Economics*, *116*(2): 607–54.

Kazianga, H., de Walque, D., and Alderman, H. (2014). 'School feeding programs, intrahousehold allocation and the nutrition of siblings: evidence from a randomized trial in rural Burkina Faso'. *Journal of Development Economics*, *106*(C): 15–34.

Kendall, S., and Bloomfield, L. (2005). 'Developing and validating a tool to measure parenting self-efficacy'. *Journal of Advanced Nursing*, *54*(2): 174–81.

Kim, J. H., Schulz, W., Zimmerman, T., and Hahlweg, K. (2018). 'Parent-Child Interactions and Child Outcomes: Evidence from Randomized Intervention'. *Labour Economics*, *54*(-): 152–71.

Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). 'Experimental Analysis of Neighborhood Effects'. *Econometrica*, *75*(1): 83–119.

Knauer, H. A., Kagawa, R. M. C., García-Guerra, A., Schnaas, L., Neufeld, L. M., and Fernald, L. C. H. (2016). 'Pathways to improved development for children living in poverty: A randomized effectiveness trial in rural Mexico'. *International Journal of Behavioral Development*, *40*(6): 492–99.

Knudsen, E. I. (2004). 'Sensitive Periods in the Development of the Brain and Behavior'. *Journal of Cognitive Neuroscience*, *16*(8): 1412–25.

La Ferrara, E. (2016). 'Mass media and social change: Can we use television to fight poverty?' *Journal of the European Economic Association*, *14*(4): 791–827.

Lu, C., Black, M. M., and Richter, L. M. (2016). 'Risk of Poor Development in Young Children in Low-Income and Middle-Income Countries: an Estimation and Analysis at the Global, Regional, and Country Level'. *Lancet Global Health*, *4*(12): 916–22.

Martinez, S., Naudeau, S., and Pereira, V. A. (2017). 'Preschool and child development under extreme poverty : evidence from a randomized experiment in rural Mozambique'. Policy Research Working Paper 8290. Washington, D.C.: The World Bank.

Matthews, J. S., Ponitz, C. C., and Morrison, F. J. (2009). 'Early gender differences in self-regulation and academic achievement.' *Journal of Educational Psychology*, *101*(3): 689.

Miller, S., and Harrison, H. (2015). 'A cluster randomised controlled trial and process evaluation of the early years DELTA parenting programme'. *International Journal of Educational Research*, *74*(-): 49–60.

Miller-Heyl, J., MacPhee, D., and Fritz, J. J. (1998). 'DARE to be you: A family-support, early prevention program'. *Journal of Primary Prevention*, *18*(3): 257–85.

Mondell, S., and Tyler, F. B. (1981). 'Parental competence and styles of problem solving/play behavior with children.' *Developmental Psychology*, *17*(1): 73.

NISR. (2018a). *Labour Force Survey Trends, February 2018.* Kigali: National Institute of Statistics of Rwanda.

NISR. (2018b). *Rwand Poverty Profile, 2016/17.* Kigali: National Institute of Statistics of Rwanda.

NISR, MOH, and ICF International. (2015a). 'Demographic and Health Survey 2014/2015'. District Profile Chart Book, Western Province. National Institute of Statistics of Rwanda (NISR), Ministry of Health of Rwanda (MOH), and ICF International.

NISR, MOH, and ICF International. (2015b). *Rwanda Demographic and Health Survey 2014/15.* Rockville, Maryland: National Institute of Statistics of Rwanda (NISR), Ministry of Health of Rwanda (MOH), and ICF International.

Ozler, B., Fernald, L. C., Kariger, P., McConnell, C., Neuman, M., and Fraga, E. (2018). 'Combining Pre-School Teacher Training with Parenting Education: A Cluster-Randomized Controlled Trial'. *Journal of Development Economics*, *133*(C): 448–67. doi: 10.1016/j.jdeveco.2018.04.004

Paluck, E. L. (2009). 'Reducing intergroup prejudice and conflict using the media: a field experiment in Rwanda.' *Journal of personality and social psychology*, *96*(3): 574.

Ray, D. (2006). 'Aspirations, poverty, and economic change'. *Understanding poverty*, *1*(-): 409–21.

Romano, J. P., and Wolf, M. (2005). 'Stepwise multiple testing as formalized data snooping'. *Econometrica*, *73*(4): 1237–82.

Roodman, D., Nielsen, M. Ø., MacKinnon, J. G., and Webb, M. D. (2019). 'Fast and Wild: Bootstrap Inference in Stata using Boottest'. *Stata Journal*, *19*(1): 4–60.

Sandner, M., and Jungmann, T. (2017). 'Gender-specific effects of early childhood intervention: evidence from a randomized controlled trial'. *Labour Economics*, *45*(-): 59–78.

Schneewind, K. A., and Pfeiffer, P. (1995). 'Impact of family processes on control beliefs'. In A. Bandura (ed.), (pp. 114–48). Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511527692.006

Shepherd, A., Scott, L., Mariotti, C., Kessy, F., Gaiha, R., da Corta, L., Hanifnia, K., Kaicker, N., Lenhardt, A., Lwanga-Ntale, C., Sen, B., Sijapati, B., Strawson, T., Thapa, G., Underhill, H., and Wild, L. (2014). *The Chronic Poverty Report 2014-2015: The road to zero extreme poverty.* London: Overseas Development Institute.

Skellern, C., Rogers, Y., and O'callaghan, M. (2001). 'A parent-completed developmental questionnaire: follow up of ex-premature infants'. *Journal of paediatrics and child health*, *37*(2): 125–29.

Spoth, R., Redmond, C., Haggerty, K., and Ward, T. (1995). 'A controlled parenting skills outcome study examining individual difference and attendance effects'. *Journal of Marriage and the Family*, *52*(2): 449–64.

Squires, J., Potter, L., and Bricker, D. (1995). *The ASQ user's guide for the Ages & Stages Questionnaires: A parent-completed, child-monitoring system.* Baltimore, MD: Paul H. Brookes Publishing.

Todd, P. E., and Wolpin, K. I. (2007). 'The production of cognitive achievement in children: Home, school, and racial test score gaps'. *Journal of Human capital*, *1*(1): 91–136.

Tomopoulos, S., Dreyer, B. P., Tamis-LeMonda, C., Flynn, V., Rovira, I., Tineo, W., and Mendelsohn, A. L. (2006). 'Books, toys, parent-child interaction, and development in young Latino children'. *Ambulatory Pediatrics*, *6*(2): 72–8.

Tucker, S., Gross, D., Fogg, L., Delaney, K., and Lapporte, R. (1998). 'The long-term efficacy of a behavioral parent training intervention for families with 2-year-olds'. *Research in Nursing & Health*, *21*(3): 199–210.

Ulfsdotter, M., Enebrink, P., and Lindberg, L. (2014). 'Effectiveness of a universal health-promoting parenting program: a randomized waitlist-controlled trial of All Children in Focus'. *BMC public health*, *14*(1): 1083.

UNDP. (2018). *Human Development Indices and Indicators: 2018 Statistical Update.* New York City: United Nations Development Programme.

Vyas, S., and Kumaranayake, L. (2006). 'Constructing socio-economic status indices: how to use principal components analysis'. *Health policy and planning*, *21*(6): 459–68.

Walker, S. P., Chang, S. M., Powell, C. A., and Grantham-McGregor, S. M. (2005). 'Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted Jamaican children: prospective cohort study'. *Lancet*, *366*(9499): 1804–07.

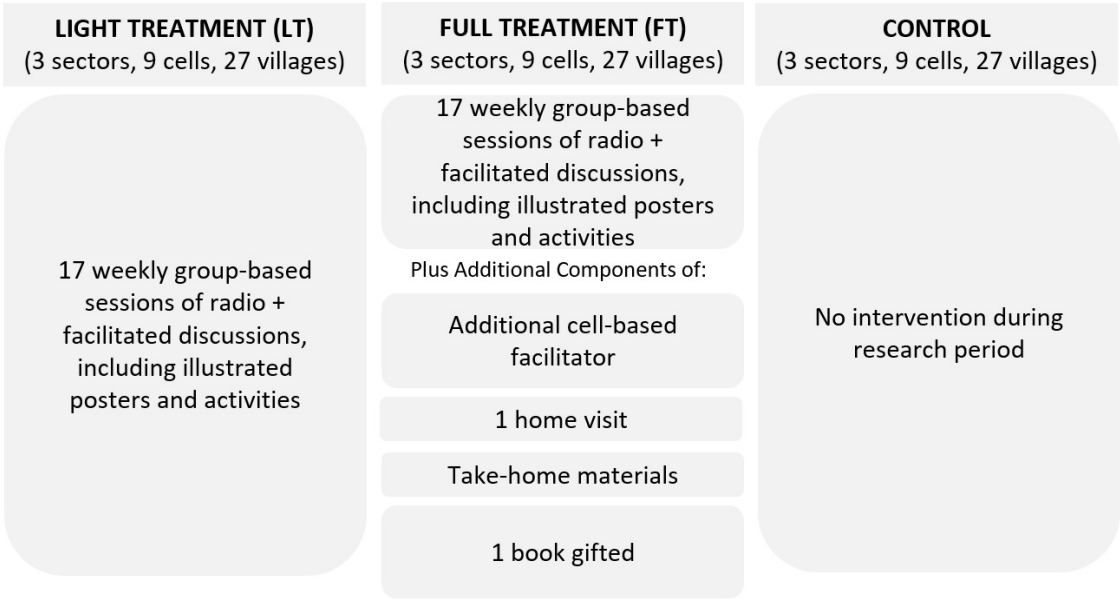
Walker, S. P., Wachs, T. D., Grantham-McGregor, S., Black, M. M., Nelson, C. A., Huffman, S. L., Baker-Henningham, H., Chang, S. M., Hamadani, J. D., Lozoff, B., et al. (2011). 'Inequality in Early Childhood: Risk and Protective Factors for Early Child Development'. *Lancet*, *378*(9799): 1325–38.

Warrinnier, N., Rozelle, S., Attanasio, O., Sylvia, S., Luo, R., Yue, A., and Medina, A. (2018). ' From Quantity to Quality: Delivering a Home-based Parenting Intervention through China's Family Planning Cadres'. LICOS Discussion Paper 402. Leuven: LICOS - Centre for Institutions and Economic Performance.

Yousafzai, A. K., Rasheed, M. A., Rizvi, A., Armstrong, R., and Bhutta, Z. A. (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'. *Lancet*, *384*(9950): 1282–93.

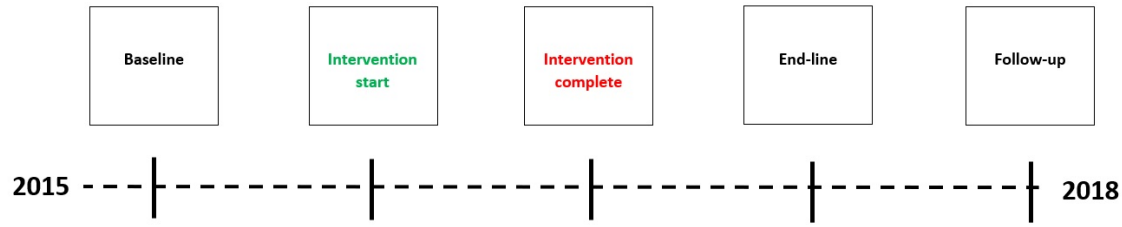
Figures

Figure 1: Experimental design



Source: authors' illustration.

Figure 2: Timeline

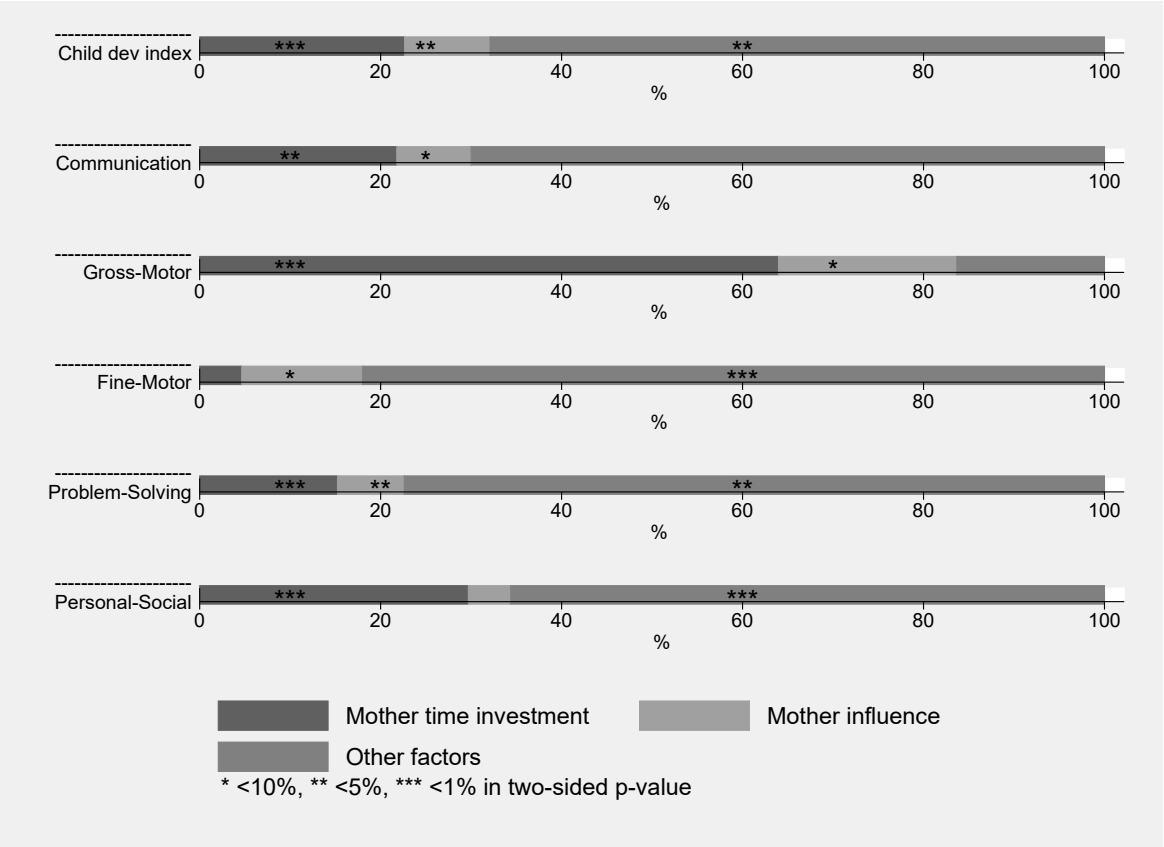


	August 2015	November 2015	April 2016	September 2016	May 2018
Child age:*	6	9	13	18	36

**Minimum age, expressed in months, of children participating to the First Steps programme. Maximum age is the age reported + 18 months.*

Source: authors' illustration.

Figure 3: Mediation analysis—short term



Notes: the figure shows results from a linear mediation analysis. The equation and the estimation process are described in Appendix C.6. Each bar represents the total treatment effect normalized to 100 per cent. The mediators displayed in each bar are mother time investment, mother influence, and other unobserved factors that include any change that is not captured by the measured parenting outcomes changes.

Source: authors' calculations based on data from end-line survey 2016.

Tables

Table 1: Child development—short term

	(1) Communication	(2) Gross motor	(3) Fine motor	(4) Problem solving	(5) Personal social	(6) Child development index
Control group - base						
Light treatment (LT)	0.316*	0.255**	0.287***	0.339**	0.361***	0.292**
	(0.145)	(0.077)	(0.076)	(0.120)	(0.065)	(0.087)
Full treatment (FT)	0.455**	0.163	0.395***	0.475***	0.472***	0.379***
	(0.143)	(0.121)	(0.076)	(0.132)	(0.068)	(0.094)
WILD p-values LT	0.055	0.038	0.017	0.046	0.016	0.022
WILD p-values FT	0.024	0.273	0.011	0.032	0.012	0.014
Romano-Wolf p-values LT	0.030	0.003	0.003	0.015	0.000	
Romano-Wolf p-values FT	0.002	0.182	0.000	0.001	0.000	
t-test LT = FT						
p-value	0.000	0.358	0.040	0.028	0.014	0.041
Observations	1428	1428	1428	1428	1428	1428
R^2	0.205	0.081	0.096	0.082	0.100	0.168

Notes: the table presents the treatment effects on child development outcomes. The sample includes children surveyed in the end-line survey (2016). All estimates show results from OLS regressions based on equation (1). All regressions include the following controls: baseline values of the outcomes variables, child gender, child age, the total number of children in the household, the caregiver's age, the caregiver's education level (defined as a binary variable equal to 1 if the caregiver has at least primary education and 0 otherwise), the caregiver's marital status (defined as a binary variable equal to 1 if the caregiver is married or cohabiting and 0 otherwise), and the asset index. This is equal to the first principal component of the following variables: floor materials of the house, roof materials of the house, main source of drinking water, and whether the house of the respondent is owned or rented, as described in Appendix A.4. All regressions include sampling weights. The dependent variables in columns 1 to 5 include standardized z-scores of the five ASQ dimensions calculated by subtracting the control group mean and dividing by the control group standard deviation in each survey wave. The dependent variable in column 6 is the child development index calculated by taking the average of the five ASQ z-scores. A full description of the construction of the outcomes is in Section 3 and in Appendix A.3. Light treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the light treatment group. Full treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the full treatment group.
*p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution are reported below the estimates (Cameron et al. 2008; Roodman et al. 2019). Two tailed p-values from a 5,000 replications Romano Wolf step-down procedure (Romano and Wolf 2005; Clarke et al. 2019) are shown below the estimates. A t-test of LT = FT is presented with the statistical significance of the test expressed in p-value. Observations and the R squared are presented at the bottom of the table.

Source: authors' calculations based on data from end-line survey (2016).

Table 2: Mother time investment—short term

	(1) Learning	(2) Positive discipline	(3) Negative discipline	(4) Mother time investment index
Control group - base				
Light treatment (LT)	0.597***	0.358***	0.218*	0.473***
	(0.029)	(0.021)	(0.095)	(0.022)
Full treatment (FT)	0.766***	0.522***	0.250**	0.623***
	(0.051)	(0.020)	(0.107)	(0.038)
WILD p-values LT	0.002	0.003	0.132	0.001
WILD p-values FT	0.005	0.001	0.101	0.004
Romano-Wolf p-values LT	0.000	0.000	0.011	
Romano-Wolf p-values FT	0.000	0.000	0.010	
t-test LT = FT				
p-value	0.006	0.000	0.688	0.002
Observations	1299	1299	1299	1299
R^2	0.278	0.119	0.031	0.269

Notes: the table presents the treatment effects on parenting outcomes. The sample includes caregivers surveyed in the end-line survey (2016). All estimates show results from OLS regressions based on equation (1). All regressions include control variables as defined in Table 1. All regressions include sampling weights. The dependent variables in columns 1 to 3 include standardized z-scores of the three HOME-SF parents time investment dimensions self-reported by the mother of the child, calculated by subtracting the control group mean and dividing by the control group standard deviation in each survey wave. The dependent variable in column 4 is the mother time investment index calculated by taking the average of the three HOME-SF z-scores. A full description of the construction of the outcomes is in Section 3 and in Appendix A.3. Light treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the light treatment group. Full treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the full treatment group.
*p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution are reported below the estimates (Cameron et al. 2008; Roodman et al. 2019). Two tailed p-values from a 5,000 replications Romano Wolf step-down procedure (Romano and Wolf 2005; Clarke et al. 2019) are shown below the estimates. A t-test of LT = FT is presented with the statistical significance of the test expressed in p-value. Observations and the R squared are presented at the bottom of the table.

Source: authors' calculations based on data from end-line survey (2016).

Table 3: Mother influence—short term

	(1) Learning	(2) Development	(3) Nutrition	(4) Care	(5) Discipline	(6) Health	(7) Mother influence index
Control group - base							
Light treatment (LT)	0.536***	0.459***	0.434***	0.430***	0.386***	0.246***	0.417***
	(0.058)	(0.060)	(0.047)	(0.034)	(0.068)	(0.067)	(0.048)
Full treatment (FT)	0.719***	0.645***	0.536***	0.620***	0.560***	0.495***	0.596***
	(0.055)	(0.070)	(0.083)	(0.071)	(0.070)	(0.069)	(0.061)
WILD p-values LT	0.011	0.009	0.008	0.008	0.012	0.021	0.009
WILD p-values FT	0.008	0.006	0.011	0.010	0.011	0.009	0.009
Romano-Wolf p-values LT	0.000	0.000	0.000	0.000	0.000	0.020	
Romano-Wolf p-values FT	0.000	0.000	0.000	0.000	0.000	0.000	
t-test LT = FT							
p-value	0.001	0.008	0.216	0.022	0.004	0.020	0.009
Observations	1300	1300	1300	1300	1300	1300	1300
R^2	0.122	0.082	0.070	0.081	0.062	0.056	0.105

Notes: the table presents the treatment effects on parenting outcomes. The sample includes caregivers surveyed in the end-line survey (2016). All estimates show results from OLS regressions based on equation (1). All regressions include control variables as defined in Table 1. All regressions include sampling weights. All regressions include sampling weights. The dependent variables in columns 1 to 6 include standardized z-scores of the six HOME-SF parental influence dimensions self-reported by the mother of the child, calculated by subtracting the control group mean and dividing by the control group standard deviation in each survey wave. The dependent variable in column 7 is the mother influence index calculated by taking the average of the six HOME-SF z-scores. A full description of the construction of the outcomes is in Section 3 and in Appendix A.3. Light treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the light treatment group. Full treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the full treatment group. *p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution are reported below the estimates (Cameron et al. 2008; Roodman et al. 2019). Two tailed p-values from a 5,000 replications Romano Wolf step-down procedure (Romano and Wolf 2005; Clarke et al. 2019) are shown below the estimates. A t-test of LT = FT is presented with the statistical significance of the test expressed in p-value. Observations and the R squared are presented at the bottom of the table.

Source: authors' calculations based on data from end-line survey (2016).

Table 4: Child development—medium term

	(1) Communication	(2) Gross motor	(3) Fine motor	(4) Problem solving	(5) Personal social	(6) Child development index
Control group - base						
Light treatment (LT)	0.108	-0.007	0.149	-0.046	0.153	0.051
	(0.095)	(0.102)	(0.122)	(0.057)	(0.083)	(0.077)
Full treatment (FT)	0.232**	0.223*	0.163	0.126*	0.225**	0.181**
	(0.095)	(0.104)	(0.110)	(0.055)	(0.091)	(0.075)
WILD p-values LT	0.373	0.955	0.435	0.478	0.152	0.588
WILD p-values FT	0.089	0.108	0.237	0.125	0.088	0.093
Romano-Wolf p-values LT	0.451	0.969	0.448	0.620	0.168	
Romano-Wolf p-values FT	0.012	0.041	0.130	0.024	0.012	
t-test LT = FT						
p-value	0.082	0.013	0.868	0.035	0.473	0.069
Observations	1256	1256	1256	1256	1256	1256
R^2	0.037	0.041	0.093	0.146	0.028	0.078

Notes: the table presents the treatment effects on child development outcomes. The sample includes children surveyed in follow-up survey (2018). All estimates show results from OLS regressions based on equation (1). All regressions include control variables as defined in Table 1. All regressions include sampling weights. The dependent variables in columns 1 to 5 include standardized z-scores of the five ASQ dimensions calculated by subtracting the control group mean and dividing by the control group standard deviation in each survey wave. The dependent variable in column 6 is the child development index calculated by taking the average of the five ASQ z-scores. A full description of the construction of the outcomes is in Section 3 and in Appendix A.3. Light treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the light treatment group. Full treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the full treatment group.

*p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution are reported below the estimates (Cameron et al. 2008; Roodman et al. 2019). Two tailed p-values from a 5,000 replications Romano Wolf step-down procedure (Romano and Wolf 2005; Clarke et al. 2019) are shown below the estimates. A t-test of LT = FT is presented with the statistical significance of the test expressed in p-value. Observations and the R squared are presented at the bottom of the table.

Source: authors' calcuations based on data from follow-up survey (2018).

Table 5: Mother time investment—medium term

	(1) Learning	(2) Positive discipline	(3) Negative discipline	(4) Mother time investment index
Control group - base				
Light treatment (LT)	0.264***	0.181***	0.105	0.211***
	(0.035)	(0.021)	(0.075)	(0.010)
Full treatment (FT)	0.309***	0.126***	0.080*	0.210***
	(0.055)	(0.036)	(0.041)	(0.030)
WILD p-values LT	0.009	0.007	0.287	0.001
WILD p-values FT	0.012	0.039	0.134	0.013
Romano-Wolf p-values LT	0.000	0.000	0.170	
Romano-Wolf p-values FT	0.000	0.014	0.029	
t-test LT = FT				
p-value	0.482	0.139	0.773	0.975
Observations	1103	1103	1103	1103
R^2	0.080	0.032	0.019	0.079

Notes: the table presents the treatment effects on parenting outcomes. The sample includes caregivers surveyed in the follow-up survey (2018). All estimates show results from OLS regressions based on equation (1). All regressions include control variables as defined in Table 1. All regressions include sampling weights. The dependent variables in columns 1 to 3 include standardized z-scores of the three HOME-SF parents time investment dimensions self-reported by the mother of the child, calculated by subtracting the control group mean and dividing by the control group standard deviation in each survey wave. The dependent variable in column 4 is the mother time investment index calculated by taking the average of the three HOME-SF z-scores. A full description of the construction of the outcomes is in Section 3 and in Appendix A.3. Light treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the light treatment group. Full treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the full treatment group.
*p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution are reported below the estimates (Cameron et al. 2008; Roodman et al. 2019). Two tailed p-values from a 5,000 replications Romano Wolf step-down procedure (Romano and Wolf 2005; Clarke et al. 2019) are shown below the estimates. A t-test of LT = FT is presented with the statistical significance of the test expressed in p-value. Observations and the R squared are presented at the bottom of the table.

Source: authors' calculations based on data from follow-up survey (2018).

Table 6: Mother self efficacy—medium term

	(1) Emotion	(2) Play	(3) Empathy	(4) Control	(5) Discipline	(6) Pressures	(7) Self-acceptance	(8) Learning	(9) Mother self-efficacy index
Control group - base									
Light treatment (LT)	0.150	0.183*	0.169*	0.090	-0.005	-0.100*	0.055	-0.021	0.065
	(0.090)	(0.086)	(0.083)	(0.073)	(0.063)	(0.046)	(0.031)	(0.063)	(0.055)
Full treatment (FT)	0.169*	0.166*	0.201*	0.151**	0.138**	0.102***	0.023	0.112*	0.133*
	(0.090)	(0.077)	(0.106)	(0.061)	(0.057)	(0.020)	(0.085)	(0.052)	(0.060)
WILD p-values LT	0.207	0.056	0.107	0.401	0.951	0.055	0.150	0.775	0.378
WILD p-values FT	0.134	0.088	0.132	0.047	0.094	0.011	0.782	0.093	0.090
Romano-Wolf p-values LT	0.232	0.109	0.148	0.380	0.934	0.108	0.209	0.874	
Romano-Wolf p-values FT	0.132	0.073	0.132	0.019	0.021	0.001	0.768	0.072	
t-test LT = FT									
p-value	0.722	0.734	0.654	0.165	0.060	0.001	0.648	0.039	0.117
Observations	1105	1105	1105	1105	1105	1105	1105	1105	1105
R^2	0.014	0.021	0.018	0.026	0.024	0.023	0.005	0.009	0.019

Notes: the table presents the treatment effects on parenting outcomes. The sample includes caregivers surveyed in the follow-up survey (2018). All estimates show results from OLS regressions based on equation (1). All regressions include control variables as defined in Table 1. All regressions include sampling weights. The dependent variables in columns 1 to 8 include standardized z-scores of the eight TOPSE parental self-efficacy dimensions self-reported by the mother of the child, calculated by subtracting the control group mean and dividing by the control group standard deviation in each survey wave. The dependent variable in column 9 is the mother self-efficacy index calculated by taking the average of the eight TOPSE z-scores. A full description of the construction of the outcomes is in Section 3 and in Appendix A.3. Light treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the light treatment group. Full treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the full treatment group.
*p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution are reported below the estimates (Cameron et al. 2008; Roodman et al. 2019). Two tailed p-values from a 5,000 replications Romano Wolf step-down procedure (Romano and Wolf 2005; Clarke et al. 2019) are shown below the estimates. A t-test of LT = FT is presented with the statistical significance of the test expressed in p-value. Observations and the R squared are presented at the bottom of the table.

Source: authors' calculations based on data from the follow-up survey (2018).

Table 7: Mother attitudes—medium term

	Child health (1)	Child behaviour (2)	Gender roles in the HH (3)	Child development (4)	Mother attitudes index (5)
Control group - base					
Light treatment (LT)	0.052	0.111*	0.104**	0.018	0.071**
	(0.038)	(0.051)	(0.042)	(0.015)	(0.025)
Full treatment (FT)	0.001	0.042	0.168**	-0.066	0.036
	(0.051)	(0.074)	(0.052)	(0.056)	(0.034)
WILD p-values LT	0.303	0.101	0.015	0.325	0.029
WILD p-values FT	0.978	0.659	0.040	0.421	0.403
Romano-Wolf p-values LT	0.226	0.066	0.039	0.226	
Romano-Wolf p-values FT	0.974	0.792	0.000	0.552	
t-test LT = FT					
p-value	0.123	0.334	0.222	0.142	0.307
Observations	1105	1105	1105	1105	1105
R^2	0.006	0.038	0.063	0.010	0.036

Notes: the table presents the treatment effects on parenting outcomes. The sample includes caregivers surveyed in the follow-up survey (2018). All estimates show results from OLS regressions based on equation (1). All regressions include control variables as defined in Table 1. All regressions include sampling weights. The dependent variables in columns 1 to 4 include standardized z-scores of the four HOME-SF parental attitudes dimensions self-reported by the mother of the child, calculated by subtracting the control group mean and dividing by the control group standard deviation in each survey wave. The dependent variable in column 5 is the mother attitudes index calculated by taking the average of the four HOME-SF z-scores. A full description of the construction of the outcomes is in Section 3 and in Appendix A.3. Light treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the light treatment group. Full treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the full treatment group.
*p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution are reported below the estimates (Cameron et al. 2008; Roodman et al. 2019). Two tailed p-values from a 5,000 replications Romano Wolf step-down procedure (Romano and Wolf 2005; Clarke et al. 2019) are shown below the estimates. A t-test of LT = FT is presented with the statistical significance of the test expressed in p-value. Observations and the R squared are presented at the bottom of the table.

Source: authors' calculations based on data from the follow-up survey (2018).

Table 8: Mother locus of control, aspiration gap, and aspirations for the child

		Aspiration gap		Aspiration for child	
	Internal Locus of Control	Become like role model	Change job	University degree	Marrying $\geq$ 22
	(1)	(2)	(3)	(4)	(5)
Control group - base					
Light treatment (LT)	0.257**	0.117*	0.027	0.108	0.095
	(0.078)	(0.060)	(0.086)	(0.060)	(0.106)
Full treatment (FT)	0.227**	0.177***	0.188***	0.148**	0.165
	(0.085)	(0.021)	(0.041)	(0.056)	(0.105)
WILD p-values LT	0.039	0.130	0.710	0.144	0.491
WILD p-values FT	0.069	0.009	0.012	0.101	0.243
t-test LT = FT					
p-value	0.776	0.310	0.117	0.648	0.243
Observations	1105	1108	1105	1105	1105
R^2	0.041	0.015	0.038	0.022	0.057

Notes: the table presents the treatment effects on parenting outcomes. The sample includes caregivers surveyed in the follow-up survey (2018). All estimates show results from OLS regressions based on equation (1). All regressions include control variables as defined in Table 1. All regressions include sampling weights. The dependent variable in column 1 is the locus of control defined as a dummy equal to 1 if the mother thinks that she is responsible for her future and 0 if she thinks it depends on luck. In columns 2 and 3 the dependent variable is the aspiration gap defined as a dummy equal to 1 if the caregiver believes that she can become as her role model within 5 years (column 2) and equal to 1 if the caregiver would like to change her job for a better life (column 3). In columns 4 and 5 the dependent variable is an indicator of the aspiration a mother has for her child defined as a dummy equal to 1 if she wishes her child to have a university degree and 0 if she wishes for her child a lower level of education (column 4), and equal to 1 if she wishes her child to marry at an age above 22 years old, and 0 if she wishes her child to marry at a younger age (below or at 22 years old). These outcomes are calculated by subtracting the control group mean and dividing by the control group standard deviation in each survey wave. A full description of the construction of the outcomes is in Section 3 and in Appendix A.3. Light treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the light treatment group. Full treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the full treatment group.
*p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution are reported below the estimates (Cameron et al. 2008; Roodman et al. 2019). Two tailed p-values from a 5,000 replications Romano Wolf step-down procedure (Romano and Wolf 2005; Clarke et al. 2019) are shown below the estimates. A t-test of LT = FT is presented with the statistical significance of the test expressed in p-value. Observations and the R squared are presented at the bottom of the table.

Source: authors' calculations based on data from the follow-up survey (2018).

Table 9: Heterogeneous treatment effect—short term

	(1) LT	(2) FT	(3) LTXBaseline characteristic	(4) FTXBaseline characteristic	(5) Baseline characteristic	(6) t-test (3) = (4)	(7) Obs
Panel A: *Heterogeneity on child development*							
Child is a girl	0.286**	0.380**	0.017	-0.013	0.003	0.770	1428
Child age $\geq$ 12 months	0.264**	0.417**	0.045	-0.065	0.201*	0.015	1428
Asset index $\geq$ median	0.135	0.273	0.201	0.123	-0.181	0.500	1428
Mother completed primary school	0.339**	0.411**	-0.105	-0.102	0.187	0.980	1428
Father completed primary school	0.339***	0.422**	-0.103	-0.122	0.183	0.706	1428
Panel B: *Heterogeneity on mother time investment*							
Child is a girl	0.532***	0.651***	-0.111	-0.047	0.078	0.120	1299
Child age $\geq$ 12 months	0.417**	0.531**	0.085	0.144**	-0.103	0.152	1299
Asset index $\geq$ median	0.488***	0.603***	-0.022	0.027	0.004	0.408	1299
Mother completed primary school	0.560***	0.665**	-0.195*	-0.111	0.151	0.120	1299
Father completed primary school	0.486***	0.668***	-0.041	-0.117	0.055	0.251	1299
Panel C: *Heterogeneity on mother influence*							
Child is a girl	0.407***	0.599***	0.020	-0.008	-0.024	0.748	1300
Child age $\geq$ 12 months	0.334*	0.495**	0.129	0.156	-0.041	0.637	1300
Asset index $\geq$ median	0.507**	0.779***	-0.108	-0.226**	0.072	0.016	1300
Mother completed primary school	0.532***	0.646***	-0.254***	-0.139	0.239*	0.325	1300
Father completed primary school	0.496***	0.683***	-0.191*	-0.233**	0.155*	0.566	1300

Notes: the table presents the heterogeneous treatment effects on child development and parenting outcomes. The sample includes children and caregivers surveyed in the end-line survey (2016). Each row shows results from an OLS regression based on equation (1) and include also an interaction term between each treatment dummy (LT and FT) and the following variables: child gender, child age, the asset index, and the caregiver's education level defined as a binary variable equal to 1 if the caregiver (both mother and father) has at least primary education and 0 otherwise. In addition, all regressions control for the dependent variable at baseline, the total number of children in the household, the caregiver's age, and the caregiver's marital status. The definition of all control variables is provided in Appendix A.4. All regressions include sampling weights. The dependent variables in Panels A, B, and C are respectively the child development aggregate index, the mother time investment aggregate index, and the mother influence aggregate index. A detailed definition of the dependent variables is provided in Section 3 and in the Appendix A.3. Light treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the light treatment group. Full treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the full treatment group.
*p <10%, **p <5%, ***p <1% reported as WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution (Cameron et al. 2008; Roodman et al. 2019). Columns 1 and 2 reports the LT and FT effects. Column 3 reports coefficients on the interaction between the LT dummy and the baseline variable. Column 4 reports coefficients on the interaction between the FT dummy and the baseline variable. Column 5 reports coefficients on the baseline variable. Column 6 reports (WILD) p-values from the t-test on the difference between the coefficients in columns 3 and 4. Column 7 reports the number of observations.

Source: authors' calculations based on data from the end-line survey (2016).

Table 10: Heterogeneity on child development at baseline—short term

	(1) Child development	(2) Mother time investment	(3) Mother influence
Base group: Control group			
Light treatment (LT)	0.353***	0.497***	0.529***
	(0.104)	(0.061)	(0.073)
Full treatment (FT)	0.357**	0.642***	0.493***
	(0.139)	(0.033)	(0.107)
Light treatment (LT) × Child development baseline: 25-50th	-0.023	0.051	-0.079
	(0.085)	(0.047)	(0.112)
Light treatment (LT) × Child development baseline: 50-75th	-0.038	-0.055	-0.212*
	(0.048)	(0.095)	(0.095)
Light treatment (LT) × Child development baseline: above 75th	-0.135**	-0.171*	-0.251***
	(0.049)	(0.086)	(0.074)
Full treatment (FT) × Child development baseline: 25-50th	0.067	0.059*	0.149
	(0.091)	(0.028)	(0.088)
Full treatment (FT) × Child development baseline: 50-75th	-0.000	-0.092*	-0.019
	(0.100)	(0.049)	(0.170)
Full treatment (FT) × Child development baseline: above 75th	-0.002	-0.099**	0.140
	(0.068)	(0.034)	(0.094)
Child development at baseline: 25-50th	0.208***	0.039***	0.118***
	(0.047)	(0.010)	(0.034)
Child development at baseline: 50-75th	0.335***	0.153***	0.186***
	(0.027)	(0.035)	(0.024)
Child development at baseline: above 75th	0.460***	0.222***	0.268***
	(0.039)	(0.022)	(0.053)
WILD p-values LT X Child development at baseline 25-50th	0.831	0.375	0.616
WILD p-values LT X Child development at baseline 50-75th	0.521	0.616	0.085
WILD p-values LT X Child development at baseline above 75th	0.024	0.186	0.024
WILD p-values FT X Child development at baseline 25-50th	0.593	0.129	0.256
WILD p-values FT X Child development at baseline 50-75th	0.997	0.201	0.889
WILD p-values FT X Child development at baseline above 75th	0.978	0.012	0.250
Observations	1428	1298	1299
R^2	0.159	0.278	0.117

Notes: the table presents the heterogeneous treatment effects on child development and parenting outcomes. The sample includes children and caregivers surveyed in the end-line survey (2016). Each column shows results from an OLS regression based on equation (1), and also includes an interaction term between each treatment dummy (LT and FT) and the child development aggregate index at baseline. This is calculated by taking the average of the five ASQ z-scores components at baseline. The variable is then defined as follows: equal to 0 if the child development index at baseline is lower than the top 25th percentile, 1 if the child development index at baseline is between the 25th and 50th percentile, 2 if the child development index at baseline is between the 50th and 75th percentile; and 3 if the child development index at baseline is above the 75th percentile. All regressions control for the dependent variable at baseline, child gender, child age, the total number of children in the household, the caregiver's age, the caregiver's marital status. The definition of all control variables is provided in Appendix A.4. All regressions include sampling weights. The dependent variable in column 1 is the child development aggregate index, in column 2 the mother time investment aggregate index, and in column 3 the mother influence aggregate index. A full description of the construction of the outcomes is in Section 3 and in Appendix A.3. Light treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the light treatment group. Full treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the full treatment group.
*p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution are reported below the estimates (Cameron et al. 2008; Roodman et al. 2019).

Source: authors' calculations based on data from the end-line survey (2016).

A First appendix

A.1 First Steps programme

Sessions

Each episode was centered on the following topics: 1) respond and bond; 2) playing together; 3) talk and read; 4) feeding with care; 5) calming and soothing; 6) the importance of setting up a routine; 7) positive discipline; 8) materials at home for play; 9) active play; 10) play and development; 11) early language and communication; 12) how to promote literacy; 13) make a book with materials found at home; 14) literacy alive at home and in the community; 15) partners for a healthy pregnancy; 16) healthy baby: prevention; 17) health clinic.

Figure A1 shows an example of the meeting location and its attendees. Figure A2 shows examples of illustrations from posters used during the meetings.

Facilitators

The activities with parents were guided by a village facilitator, called a Community Family Volunteer (CFV), drawn from a network of local women and men. CFVs were chosen by the local leaders in the implementation cells and villages. Save the Children organized an awareness meeting at the sector level and requested local leaders to provide names of potential candidates for the role of village facilitator. Criteria for selection included: (i) living in the village where the intervention was implemented, (ii) being literate, and (iii) not having any previous criminal record. Each village then organized a committee comprising the most influential individuals in the village, who were tasked with accepting or rejecting the candidates. Once selected, village facilitators were trained by Save the Children and Umuhuza for three and a half days. They also received a monthly coaching session by Umuhuza for the duration of the intervention. The training was focused on four areas (also the focus of the radio show): responsive caring, playful learning, language and communication, and healthy meetings. The CFV received RWF4,000 per month as an incentive.

On average 60 per cent of the facilitators were females and the average age was 36 years old. Furthermore, 25 per cent had primary education, 65 per cent completed at least junior secondary education and the remaining ten per cent completed vocational training.

The Full Treatment intervention arm also included a cell-based facilitator, called a Community Family Facilitator (CFF), who was involved in supporting and supervising the work of the village facilitators for their cell. Families in the FT arm received one home visit from the CFV and CFF, in addition to the group sessions. This provided an opportunity to review with the caregiver the messages learned during the group meetings with the support of take home cards, and discuss any challenge related to applying learning. The CFF got the same amount of training than the village facilitators, and received RWF4,500 per month. In each village, CFVs and CFFs were two distinguished individuals known to the families.

A.2 Data and sampling

The intervention was evaluated using a cluster-randomized controlled trial with a control group and two treatment arms implemented in the Western Province in the Ngororero district. Rwanda is divided in five provinces, 30 districts, 416 sectors, 2,148 cells, and approximately 14,837 villages. Provinces have a minimum of three districts (Kigali city) and a maximum of eight districts (Southern Province). The Western Province includes seven districts. One of these is Ngororero, which includes 13 sectors.

There are six cells in each sector, and each cells contains around four to five villages (NISR 2018a). Within the Ngororero district, nine sectors out of 13 were selected for the study. Three out of the 13 sectors in the Ngororero district were excluded because a pilot of First Steps had been implemented in these areas. One additional sector was excluded due to the lack of a market that would support local booksellers. Within each sector, three cells were randomly selected and, within each cell, three villages were randomly selected. The final sample included nine sectors, 27 cells, and 81 villages.

A.3 Tools and variables

Ages and Stages Questionnaire (ASQ)

We used the ASQ third edition, originally developed by Squires et al. (1995). The ASQ tool requires the principal caregiver to answer questions related to five child development domains. During some activities, the caregiver is required to interact with the child, while the enumerator observes and collects data from a distance. Child development is assessed along five distinct development dimensions (communication, gross motor, fine motor, problem solving, and personal-social skills), covering a total of 30 questions. Given the age-specific nature of the questions, the questionnaire is specifically tailored to different age ranges, with a different age span for each questionnaire. The questionnaire has a two-month span for the two to 24 months questionnaires; a three-month span for the 24–36 months questionnaires; and a six-month span for the 36–60 months questionnaires. For simplicity, the ASQ was administered using a six-month age span. For each child activity, the enumerator chooses between three options: i) yes; ii) sometimes; iii) not yet, which are coded with a score. 'Yes' indicates that the child can perform the skill/action, coded as ten points. 'Sometimes' indicates that the child is just beginning to perform the behavior (i.e. emerging skill), or performs the skill on occasion but not all the time, coded as five points. 'Not yet' indicates that the child is not yet performing the skill, coded as 0 point.

At the end of the assessment, a total score is produced for each skill by adding up all the activity scores achieved. Scores range between 0 and 60 for each dimension of child development. The large majority of the activities in the communication, gross motor skills, fine motor, problem solving, and half of the questions on personal social skills are observed and reported by the enumerator using the above coding system. Twelve questions out of 192 on the first four ASQ dimensions (communication, gross motor, fine motor, and problem solving) and the remaining half of the questions on personal social skills are self-reported by the caregiver.

We constructed raw scores for each dimension by adding up the scores obtained for each activity/skill in the appropriate age range. The minimum a child can achieve for each of the five dimensions is 0 while the maximum is 60. We then calculated standardized scores by subtracting the mean of the control group and dividing by the standard deviation of the control group in each survey wave. We calculated the aggregate index by taking the average of the five z-scores.

Home Observation for Measurement of the Environment - Short Form (HOME-SF)

We collected data on parenting practices using the HOME-SF questionnaire, adapted for the Rwandan context and translated in Kinyarwanda. The HOME-SF questionnaire was originally developed and later modified by Bradley and Caldwell (1977). We constructed the main parenting outcomes as follows:

Parental time investment: In the HOME-SF, the principal caregiver is asked how often they perform a certain activity with the child. The principal caregiver answers for themselves and on behalf of their partner. We divided the answers into maternal time investment when the respondent is the mother and she answers for herself, and into paternal time investment when the respondent is the father and he answers for himself. Given the structure of the sample and the fact that, in the large majority of households,

mothers are the principal caregivers, most answers are provided by the mother, who answers questions about her interactions with the child. The sample includes less than 100 observations in which the father answers questions related to his own interactions with the child. We also analysed the outcomes using the reported answers of the mother related to father time investment, and vice-versa, which allows to obtain a larger sample for father time investment. However, as it is the mother (father) reporting on her (his) partner activities, this measure might not reflect accurately the activities performed by the partner and we are concerned this larger sample may include large measurement error. Results using this larger sample were nonetheless similar to those obtained using the answers that pertained directly to the respondent (which are reported in the paper) and are available upon request.

The respondent reported on the frequency with which they perform a certain activity with the child. Possible answers were as follows: (0) not at all, (1) rarely, (2) a few times each month, (3) a few times each week, (4) once per day, (5) more than once per day. Two questions, (i) do you shout at your child and (ii) do you hit your child, were reverse coded.

At follow-up (May 2018), we did not ask about the frequency of interactions, but asked only whether the parent performed a certain activity or not with the child. This was done in order to reduce the length of the questionnaire due to time and budget constraints. At follow-up, we asked about 15 activities instead of the full set of 18 activities asked at baseline and end-line. This was done because some activities were similar and could be aggregated (e.g. play and play inside with toys). We also rephrased slightly some activities, but we made sure that the meaning was kept the same. Two activities—making books with traditional materials and take away something child wanted—were only asked at follow-up.

For each activity, we constructed standardized scores. They were calculated by subtracting the control group mean of answers for each activity in each survey wave (baseline, end-line, and follow-up), and then dividing by the standard deviation of the control group. Thus, the standardized score for each activity has, for the control group, mean 0 and standard deviation 1. We then constructed an aggregate index by calculating the average of the z-scores of each activity.

Parental influence: Each respondent was asked how much influence they think they have over their child, along six different dimensions: (i) learning, (ii) development, (iii) nutrition, (iv) care, (v) discipline or guidance, and (vi) health care. The level of influence ranged between: (0) no influence; (1) some influence; (2) much influence; and (3) very much influence.

We constructed a standardized score for each of these six dimensions. The standardized scores were calculated by subtracting from the mean of the control group the answers obtained by each respondent in each survey wave (baseline, end-line, and follow-up), and then dividing this number by the standard deviation of the control group. Thus, the standardized score of each dimension has, for the control group, mean 0 and standard deviation 1. We created also means, for each survey wave, of the six standardized dimensions listed above, which were used in the 12-month analysis.

Additional tools and variables

Parental self-efficacy: In order to measure parental self-efficacy we used the *Tool to measure Parenting Self Efficacy* (TOPSE) developed by Kendall and Bloomfield (2005). TOPSE is based on 48 statements organized in eight sections: (1) emotion and affection; (2) play and enjoyment; (3) empathy and understanding; (4) control; (5) discipline and setting boundaries; (6) pressures; (7) self-acceptance; (8) learning and knowledge. TOPSE is the most complete parental self-efficacy questionnaire available.

Parents were invited to rate how much they agreed with each statement on a scale of 0–10, from 'completely disagree' to 'completely agree'. After piloting the questionnaire in Rwanda, and discussing with the developers of TOPSE, we decided to use the low literacy version of the tool, which includes 43 statements in simpler language, organized on a scale ranging between 1 and 5. To facilitate the administration

of the tool, as suggested by the literature (Delavande et al. 2011), we also used a scale in the format of a picture, as presented in Figure A3. The enumerators were trained to read the following statement out loud before administering this question: '*The following section is about section 1 (i.e. emotion and affection). Using the scale below, please enter in the boxes how much you agree with each statement. The scale ranges from 1 (completely disagree) to 5 (completely agree). You may use any number between 1 and 5. Please answer all statements*'. The enumerator asked, for example: 'Can you point on the scale how much do you agree from 1 to 5 with the statement *'I can show my child I love her'*?' (this statement is taken from the first question of the section 'emotion and affection'). To construct the outcome of interest, we added up responses for each of the eight sections. We then standardized these indicators by subtracting the mean and dividing by the standard deviation of the control group in each survey round. We also created an aggregate index by taking the mean of the standardized sums. This aggregate index was then analysed for the mother only, when the respondent was the mother, and for the father only, when the respondent was the father.

Parental attitudes: Parental attitudes are examined along four dimensions: child health, behavior, gender, and child development. Each answer is solicited as a dummy equal to 1 if the parent agrees, and 0 otherwise. We have reverse coded the negative activities. We constructed standardized scores for each activity. The standardized scores were calculated by subtracting the mean of the control group for each activity recorded at follow-up, and then dividing by the standard deviation of the control group. Thus, the standardized score of each activity has, for the control group, mean 0 and standard deviation 1. We calculated also an average aggregate index by taking the mean of the four standardized indices.

Locus of control: Locus of control is defined as the individual perception or belief about the underlying causes of events in their life (Dercon et al. 2012). We study this variable using a question constructed as a dummy, which takes the value of 1 if the person thinks that they are responsible for their future, and 0 if they think their future is determined by luck (i.e. they have no control over it). We calculated a standardized score for each question with respect to the control group in the relevant survey wave by subtracting the mean, and dividing by the standard deviation of the control group.

Aspirations gap: Aspiration gap is analysed using two questions. First, we constructed a binary indicator that equals 1 if the caregiver believes they can become like an a priori defined role model within the next five years. Second, we constructed a binary indicator equal to 1 if the caregiver would like to change their job in order to achieve a better life. We calculated a standardized score with respect to the control group in the relevant survey wave by subtracting the mean, and dividing by the standard deviation of the control group.

Aspirations for children: We analysed the aspirations of the caregiver towards their children using two questions. The first question asked whether the caregiver wishes their child to have a university degree. Answers were coded as 1 if the caregiver answered 'yes' and 0 otherwise. The second question asked whether the caregiver wished their child married only after they turned 22 years old. As above, answers were coded as 1 if the caregiver answered 'yes' and 0 if not. We calculated standardized scores for each statement with respect to the control group in the relevant survey wave by subtracting the mean, and dividing by the standard deviation of the control group.

A.4 Additional variable definitions

We use the following variables in the estimated specifications:

Child gender: dummy variable coded as 1 if the child is a girl and 0 otherwise.

Child age: continuous variable corresponding to the child's age in months. At baseline, all children are aged between six and 24 months old.

Number of children in the household: continuous variable indicating the number of children in the household at the time of the baseline survey.

Respondent age: continuous variable indicating the respondent's age in years.

Mother (father) highest level of education achieved: ordinal variable corresponding to the highest education achieved by the mother (father) of the child. Education levels are the following: (0) no formal education, (1) primary school, (2) secondary school, (3) higher education, (4) vocational training, and (5) ordinary level. Because the majority of respondents either do not have formal education, or achieved only primary education, we created a dummy variable coded as 1 if the mother (father) of the child has at least completed primary education, and 0 otherwise.

Marital status of the respondent: categorical variable indicating the marital status of the respondent at the time of the survey. The categories are defined as (1) single, (2) married, (3) cohabiting, (4) widow, and (5) divorced. Because the majority of respondents are married, we created a binary variable coded as 1 if the respondent is married or cohabiting, and 0 otherwise.

Asset index: variable constructed using a principal component analysis of variables, which are floor materials of the house, roof materials of the house, main source of drinking water, and whether the house of the respondent is owned or rented. Given the lack of variation in the answers over ownership of the following items: bicycle, motorcycle, car, radio, television, refrigerator, computer, and electricity access, we avoided taking into account these variables when constructing the asset index (Vyas and Kumaranayake 2006). For each floor, roof, drinking water, and home ownership typology, we grouped the categories in an asset with (i) low value, (ii) median value, and (iii) high value. This qualitative distinction is done to group similar items together from a monetary/wealth perspective. The outcome of the principal component analysis is a table of factor scores (or weights) for each of the four variable typologies. A positive score is associated with a higher socio-economic status.

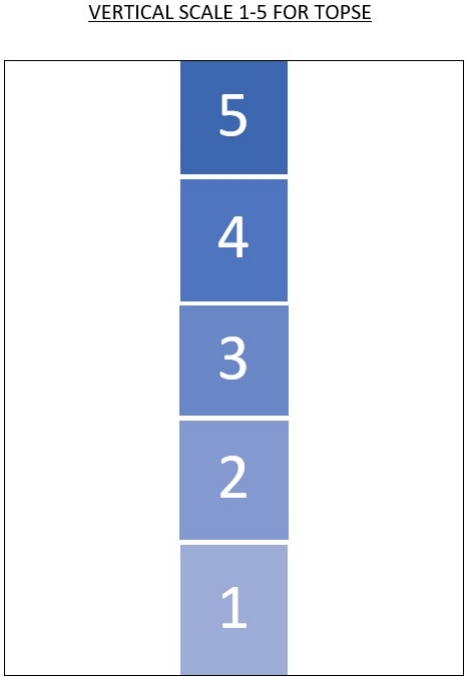
Figure A1: First Steps sessions

Source: courtesy of Save the Children Rwanda.

Figure A2: First Steps material

Source: courtesy of Save the Children Rwanda.

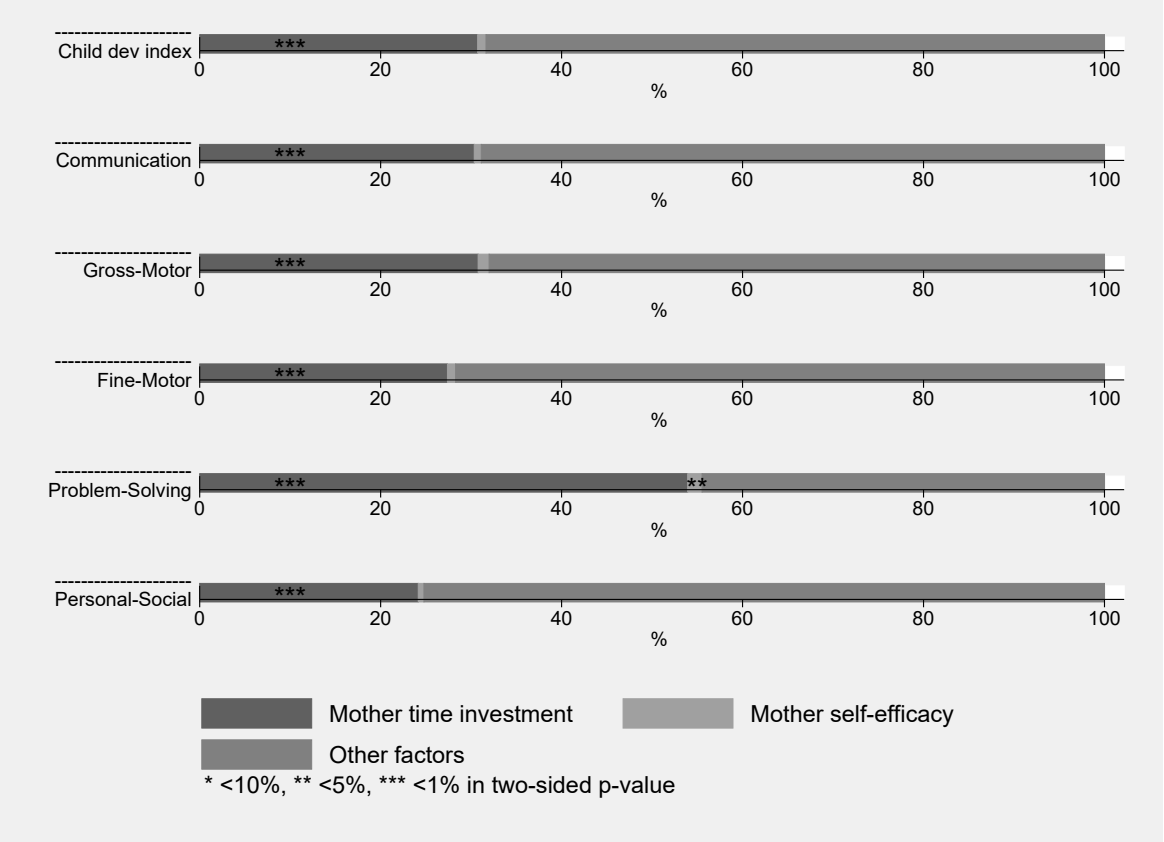
Figure A3: TOPSE low literacy scale



Source: authors' illustration.

B Additional figures and tables

Figure B1: Mediation analysis—medium term



Notes: the figure shows results from a linear mediation analysis. The equation and the estimation process are described in Appendix C.6. Each bar represents the total treatment effect normalized to 100 per cent. The mediators displayed in each bar are: mother time investment, mother self-efficacy, and other unobserved factors that include any change that is not captured by the measured parenting outcomes changes.

Source: authors' calculations based on data from the follow-up survey (2018).

Table B1: Baseline balance

	Control (CG)	Light treatment (LT)				Full treatment (FT)				LT=FT	Obs
	Mean	Mean diff. LT - CG	Unadj. pvalue	Wild pvalue	Normal. diff.	Mean diff. FT - CG	Unadj. pvalue	Wild pvalue	Normal. diff.	Wild pvalue	
Panel A: *Child outcomes and characteristics*											
ASQ communication z-score	-0.00	0.09	0.10	0.15	0.07	0.11	0.18	0.27	0.08	0.86	1613
ASQ gross motor z-score	0.00	0.28	0.00	0.01	0.20	0.13	0.36	0.43	0.09	0.46	1613
ASQ fine motor z-score	-0.00	0.28	0.00	0.00	0.21	0.24	0.03	0.10	0.17	0.73	1613
ASQ problem solving z-score	-0.00	0.32	0.00	0.01	0.24	0.19	0.17	0.28	0.13	0.51	1613
ASQ personal social z-score	-0.00	0.27	0.00	0.01	0.20	0.16	0.14	0.19	0.12	0.35	1613
Child development index	-0.00	0.25	0.00	0.01	0.24	0.17	0.12	0.27	0.15	0.58	1613
Child is a girl	0.57	-0.08	0.00	0.01	-0.11	-0.06	0.01	0.01	-0.08	0.16	1614
Child age in months	14.49	0.48	0.32	0.39	0.06	-0.25	0.58	0.68	-0.03	0.09	1614
Panel B: *Parents outcomes and characteristics*											
Mother time investment	-0.00	0.03	0.32	0.41	0.03	0.10	0.05	0.13	0.15	0.30	1504
Mother influence	-0.00	-0.07	0.51	0.61	-0.07	-0.00	0.98	0.97	-0.01	0.35	1506
Father time investment	0.00	-0.14	0.23	0.28	-0.07	-0.12	0.64	0.65	0.02	0.94	86
Father influence	-0.00	-0.03	0.86	0.87	-0.07	-0.11	0.76	0.81	-0.01	0.84	86
Respondent is mother	0.93	-0.00	0.96	0.96	-0.00	0.01	0.07	0.12	0.04	0.64	1614
Respondent father	0.06	-0.01	0.64	0.67	-0.02	-0.02	0.10	0.17	-0.05	0.71	1614
Respondent age	29.55	0.99	0.01	0.04	0.10	0.50	0.20	0.30	0.05	0.39	1614
Number of children in the HH	2.96	-0.23	0.01	0.02	-0.09	0.01	0.85	0.86	0.01	0.01	1614
Family asset index (factor variable)	0.08	-0.23	0.51	0.60	-0.10	-0.01	0.96	0.97	-0.01	0.61	1614
Mother highest education in %: Primary onwards	0.38	0.17	0.02	0.04	0.25	-0.03	0.64	0.70	-0.05	0.01	1614
Father highest education in %: Primary onwards	0.40	0.09	0.18	0.29	0.13	-0.02	0.76	0.80	-0.03	0.01	1614
Respondent is married or cohabitating in %	0.90	-0.06	0.11	0.17	-0.12	0.02	0.35	0.43	0.05	0.07	1614

Notes: the sample in this table only includes observations related to children (Panel A) and caregivers (Panel B) surveyed at baseline in 2015. Column 1 reports the mean of observations in the control group at baseline from a regression that includes sample weights and robust standard errors clustered at the sector level. Column 2 reports the difference in means of the observations in the light treatment and the control group at baseline. Column 3 shows the unadjusted p-values for the null hypothesis of no difference between observations in the light treatment and the control group. Column 4 reports the p-values based on a two-tailed mean t-test for the null hypothesis of no difference between light treatment and the control group. It uses a WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution as proposed by Cameron et al. (2008), and developed by Roodman et al. (2019). Column 5 reports the normalized difference between the observations in the light treatment and the control group computed as the difference in means in treatment and control observations divided by the square root of the sum of the variances. Column 6–9 report the corresponding statistics for the full treatment group. Column 10 reports p-values of a t-test for the null hypothesis of no difference between the light treatment and full treatment groups. Column 11 shows the number of observations at baseline. All variables are described in Section 3 and Appendix A.3.

Source: authors' calculations based on data from baseline survey (2015).

Table B2: Participation in First Steps sessions

	Obs	Mean	St.dev.	Min	Max
In which sessions did parent participate?					
Child development	972	0.71	0.45	0	1
Nutrition and breastfeeding	972	0.46	0.50	0	1
Health	972	0.37	0.48	0	1
Positive discipline	972	0.37	0.48	0	1
Early Literacy	972	0.36	0.48	0	1
Early Math	972	0.18	0.39	0	1
Responsive caring	972	0.56	0.50	0	1
Play	972	0.53	0.50	0	1

Notes: the table shows summary statistics of the specific sessions participants remember having attended. The sample only includes caregivers in LT and FT groups.

Source: authors' calculations based on data from the end-line survey (2016).

Table B3: Correlates of take-up

	Number of sessions attended (1)
Child is a female	-0.103
	(0.242)
Child age in months	0.028
	(0.021)
Total number of children	-0.065
	(0.135)
Caregiver age	0.024
	(0.036)
Mother education: primary or onwards	0.512*
	(0.228)
Father education: primary or onwards	0.254
	(0.283)
Caregiver is married or cohabitating	-0.172
	(0.393)
HH asset index	0.240**
	(0.069)
Observations	952
R^2	0.014

Notes: the sample includes caregivers surveyed in the end-line survey (2016). All estimates show results from OLS regressions. The dependent variable is the number of sessions a caregiver remembers having attended, and assumes values from 1 to 17. The regression includes baseline values of child gender, child age, the total number of children in the household, the caregiver's age, the caregiver's education level, the caregiver's marital status, and the asset index. A full description of the variables is provided in Appendix A.4. The regression includes sampling weights.
*p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. Observations and the R squared are presented at the bottom of the table.

Source: authors calculations based on data from the end-line survey (2016).

Table B4: Intensity of treatment

Panel A: Short term	(1) Child development	(2) Mother time investment	(3) Mother influence
Number of sessions attended (out of 17)	0.022***	0.035***	0.036***
	(0.006)	(0.004)	(0.005)
WILD p-values N sessions attended	0.005	0.000	0.001
Observations	1427	1298	1299
R^2	0.162	0.221	0.099

Panel B: Medium term	(1) Child development	(2) Mother time investment	(3) Mother self-efficacy
Number of sessions attended (out of 17)	0.009	0.013***	0.006*
	(0.005)	(0.001)	(0.003)
WILD p-values N sessions attended	0.139	0.001	0.032
Observations	1255	1102	1104
R^2	0.073	0.071	0.016

Notes: the sample includes children and caregivers surveyed in the end-line (Panel A) and in the follow-up (Panel B) surveys. All estimates show results from OLS regressions. The main independent variable in all columns is the number of sessions a caregiver remembers having attended which assumes values from 0 to 17. The variable assumes values equal to 0 for the observations in the control group. All regressions control for baseline values of the dependent variable, child gender, child age, the total number of children in the household, the caregiver's age, the caregiver's education, the caregiver's marital status, and the asset index. All control variables are described in Appendix A.4. All regressions include sampling weights. The dependent variables are the child development index (column 1, Panel A and B), the mother time investment index (column 2, Panel A and B), the mother influence index (column 3, Panel A), and the mother self-efficacy index (column 3, Panel B). A full description of the construction of the outcomes is in Section 3 and in Appendix A.3.
*p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution are reported below the estimates (Cameron et al. 2008; Roodman et al. 2019). Observations and the R squared are presented at the bottom of the table.

Source: authors' calculations based on data from the end-line (2016) and follow-up (2018) surveys.

Table B5: Father time investment—short term

	(1) Learning	(2) Positive discipline	(3) Negative discipline	(4) Father time investment index
Control group - base				
Light treatment (LT)	0.239	0.234**	-0.154	0.177*
	(0.199)	(0.073)	(0.228)	(0.094)
Full treatment (FT)	0.801**	0.410*	-0.249	0.537**
	(0.291)	(0.196)	(0.247)	(0.216)
WILD p-values LT	0.297	0.015	0.563	0.185
WILD p-values FT	0.012	0.062	0.458	0.017
Romano-wolf p-values LT	0.365	0.021	0.498	
Romano-wolf p-values FT	0.012	0.057	0.279	
t-test LT = FT				
p-value	0.073	0.452	0.731	0.139
Observations	87	87	87	87
R^2	0.363	0.150	0.258	0.304

Notes: the table presents the treatment effects on father time investment. The sample includes fathers surveyed in the end-line survey (2016). All estimates show results from OLS regressions based on equation (1). All regressions include control variables as defined in Table 1. All regressions include sampling weights. The dependent variables in columns 1 to 3 include standardized z-scores of the three HOME-SF parents time investment dimensions self-reported by the father of the child, calculated by subtracting the control group mean and dividing by the control group standard deviation in each survey wave. The dependent variable in column 4 is the father time investment index calculated by taking the average of the three HOME-SF z-scores. A full description of the construction of the outcomes is in Section 3 and in Appendix A.3. Light treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the light treatment group. Full treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the full treatment group.
*p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution are reported below the estimates (Cameron et al. 2008; Roodman et al. 2019). Two tailed p-values from a 5,000 replications Romano Wolf step-down procedure (Romano and Wolf 2005; Clarke et al. 2019) are shown below the estimates. A t-test of LT = FT is presented with the statistical significance of the test expressed in p-value. Observations and the R squared are presented at the bottom of the table.

Source: authors' calculations based on data from the end-line survey (2016).

Table B6: Mother and father participation in First Steps

Panel A: Short term	(1) Child development	(2) Mother time investment	(3) Mother influence
Base group: No participation (Control group or non-compliers)			
Father participated alone	0.225	0.430**	0.417
	(0.180)	(0.159)	(0.330)
Mother participated alone	0.293***	0.510***	0.473***
	(0.079)	(0.047)	(0.067)
Mother and father participated together	0.431***	0.634***	0.499***
	(0.095)	(0.079)	(0.102)
WILD p-values Father participated alone	0.334	0.102	0.352
WILD p-values Mother participated alone	0.007	0.001	0.004
WILD p-values Mother and father participated together	0.005	0.001	0.003
Observations	1414	1288	1289
R^2	0.162	0.243	0.090

Panel B: Medium term	(1) Child development	(2) Mother time investment	(3) Mother self-efficacy
Base group: No participation (Control group or non-compliers)			
Father participated alone	0.266	0.045	0.496***
	(0.293)	(0.175)	(0.142)
Mother participated alone	0.103	0.199***	0.065
	(0.076)	(0.017)	(0.058)
Mother and father participated together	0.156	0.220***	0.064
	(0.117)	(0.031)	(0.069)
WILD p-values Father participated alone	0.416	0.852	0.044
WILD p-values Mother participated alone	0.272	0.001	0.361
WILD p-values Mother and father participated together	0.269	0.001	0.385
Observations	1188	1064	1066
R^2	0.075	0.081	0.017

Notes: the table reports results from a regression that looks at the correlation between who in the family participated in First Steps and child development and parenting outcomes. The sample includes children and caregivers surveyed in the end-line survey (Panel A) and in the follow-up survey (Panel B). All estimates show results from OLS regressions. All regressions include control variables as defined in Table 1 and include sampling weights. All control variables are described in Appendix A.4. All regressions include sampling weights. The dependent variables are the child development index (column 1, Panel A and B), the mother time investment index (column 2, Panel A and B), the mother influence index (column 3, Panel A) and the mother self-efficacy index (column 3, Panel B). A full description of the construction of the outcomes is in Section 3 and in Appendix A.3. The main independent variable is defined as equal to 0 if the caregiver did not participate (reference group: control group and non-compliers), equal to 1 if the father participated alone, equal to 2 if the mother participated alone, and equal to 3 if they participated together.

*p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution are reported below the estimates (Cameron et al. 2008; Roodman et al. 2019). Observations and the R squared are presented at the bottom of the table.

Source: authors' calculations based on data from the end-line (2016) and follow-up (2018) surveys.

Table B7: Mother time investment by activity—short term

	Light treatment (LT)				Full treatment (FT)				LT=FT	Obs
	β	Unadj.	WILD	Rom. Wolf	β	Unadj.	WILD	Rom. Wolf	Unadj.	
Play	0.480	0.000	0.008	0.000	0.591	0.000	0.007	0.000	0.012	1299
Sing	0.635	0.000	0.006	0.000	0.709	0.000	0.009	0.000	0.221	1299
Tell a story	0.584	0.000	0.010	0.000	0.844	0.000	0.009	0.000	0.019	1299
Play with toys	1.063	0.000	0.000	0.000	1.301	0.000	0.003	0.000	0.011	1299
Take the child out	0.532	0.000	0.011	0.000	0.640	0.000	0.010	0.000	0.084	1299
Visit relatives	0.486	0.000	0.004	0.000	0.710	0.000	0.004	0.000	0.002	1299
Teach something new	0.431	0.000	0.009	0.000	0.606	0.000	0.008	0.000	0.041	1299
Name objects	0.520	0.000	0.009	0.000	0.724	0.000	0.007	0.000	0.001	1299
Count	0.646	0.000	0.007	0.000	0.831	0.000	0.005	0.000	0.020	1299
Hug	0.401	0.000	0.000	0.000	0.522	0.000	0.005	0.000	0.031	1299
Soothe	0.270	0.001	0.010	0.001	0.500	0.000	0.009	0.000	0.000	1299
Respond verbally	0.291	0.000	0.010	0.000	0.472	0.000	0.007	0.000	0.001	1299
Praise	0.426	0.000	0.007	0.000	0.625	0.000	0.005	0.000	0.000	1299
Positive discipline	0.415	0.000	0.009	0.000	0.513	0.000	0.007	0.000	0.009	1299
Do not shout	0.019	0.885	0.897	0.852	0.060	0.623	0.632	0.603	0.617	1299
Do not hit	0.412	0.001	0.013	0.000	0.441	0.005	0.019	0.001	0.780	1299
Read books	1.312	0.000	0.006	0.000	2.068	0.000	0.004	0.000	0.001	1299
Show books	1.103	0.000	0.004	0.000	1.880	0.000	0.001	0.000	0.000	1299

Notes: the table presents estimates of treatment effects on mother time investment in each activity. The sample includes caregivers surveyed in the end-line survey (2016). All estimates show results from separate OLS regressions based on equation (1). The dependent variables are the single components of the mother time investment index defined in Section 3 and in Appendix A.3. All regressions include control variables as defined in Table 1 and sampling weights. Columns 1 and 5 report the light treatment and full treatment coefficients for each dependent variable, respectively. Columns 2 and 6 report the unadjusted two-tailed p-values. Columns 3 and 7 show the WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution (Cameron et al. 2008; Roodman et al. 2019). Columns 4 and 8 show the two tailed p-values from a 5,000 replications Romano Wolf step-down procedure (Romano and Wolf 2005; Clarke et al. 2019). Column 9 reports p-values of a t-test for the null hypothesis of no difference between the light treatment and full treatment groups. Column 10 shows the number of observations.

Source: authors' calculations based on data from the end-line survey (2016).

Table B8: Father influence—short term

	(1) Learning	(2) Development	(3) Nutrition	(4) Care	(5) Discipline	(6) Health	(7) Father influence index
Control group - base							
Light treatment (LT)	0.645***	0.504	0.335	0.226	0.564*	0.400	0.452*
	(0.185)	(0.379)	(0.356)	(0.125)	(0.284)	(0.233)	(0.225)
Full treatment (FT)	0.943**	0.669*	0.687**	0.915***	0.405*	0.504**	0.681***
	(0.300)	(0.330)	(0.275)	(0.229)	(0.201)	(0.163)	(0.194)
WILD p-values LT	0.007	0.316	0.404	0.063	0.086	0.178	0.088
WILD p-values FT	0.034	0.138	0.081	0.002	0.243	0.097	0.063
Romano-wolf p-values LT	0.029	0.235	0.311	0.154	0.146	0.168	
Romano-wolf p-values FT	0.008	0.077	0.049	0.003	0.077	0.008	
t-test LT = FT							
p-value	0.425	0.669	0.254	0.032	0.663	0.670	0.413
Observations	87	87	87	87	87	87	87
R^2	0.279	0.188	0.159	0.203	0.164	0.111	0.248

Notes: the table presents the treatment effects on father influence. All estimates show results from OLS regressions based on equation (1). All regressions include control variables as defined in Table 1 and sampling weights. The dependent variables in columns 1 to 6 include standardized z-scores of the six HOME-SF parental influence dimensions self-reported by the father of the child, calculated by subtracting the control group mean and dividing by the control group standard deviation in each survey wave. The dependent variable in column 7 is the father influence index calculated by taking the average of the six HOME-SF z-scores. A full description of the construction of the outcomes is in Section 3 and in Appendix A.3. Light treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the light treatment group. Full treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the full treatment group.
*p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution are reported below the estimates (Cameron et al. 2008; Roodman et al. 2019). Two tailed p-values from a 5,000 replications Romano Wolf step-down procedure (Romano and Wolf 2005; Clarke et al. 2019) are shown below the estimates. A t-test of LT = FT is presented with the statistical significance of the test expressed in p-value. Observations and the R squared are presented at the bottom of the table.

Source: authors' calculations based on data from the end-line survey (2016).

Table B9: Father time investment—medium term

	(1) Learning	(2) Positive discipline	(3) Negative discipline	(4) Father time investment index
Control group - base				
Light treatment (LT)	-0.083	0.084	-0.086	-0.036
	(0.201)	(0.099)	(0.144)	(0.080)
Full treatment (FT)	0.299**	0.179	-0.082	0.183**
	(0.117)	(0.123)	(0.098)	(0.056)
WILD p-values LT	0.735	0.442	0.566	0.697
WILD p-values FT	0.031	0.236	0.443	0.008
Romano-wolf p-values LT	0.776	0.680	0.776	
Romano-wolf p-values FT	0.036	0.248	0.404	
t-test LT = FT				
p-value	0.033	0.390	0.981	0.020
Observations	108	108	108	108
R^2	0.208	0.136	0.050	0.216

Notes: the table presents the treatment effects on father time investment. All estimates show results from OLS regressions based on equation (1). All regressions include control variables as defined in Table 1 and sampling weights. The dependent variables in columns 1 to 3 include standardized z-scores of the three HOME-SF parents time investment dimensions, self-reported by the father of the child, calculated by subtracting the control group mean and dividing by the control group standard deviation in each survey wave. The dependent variable in column 4 is the father time investment Index calculated by taking the average of the three HOME-SF z-scores. A full description of the construction of the outcomes is in Section 3 and in Appendix A.3. Light treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the light treatment group. Full treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the full treatment group. *p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution are reported below the estimates (Cameron et al. 2008; Roodman et al. 2019). Two tailed p-values from a 5,000 replications Romano Wolf step-down procedure (Romano and Wolf 2005; Clarke et al. 2019) are shown below the estimates. P-values of a t-test for the null hypothesis of no difference between the light treatment and full treatment groups are also reported, number of observations and the R squared are presented at the bottom of the table.

Source: authors' calculations based on data from the follow-up survey (2018).

Table B10: Mother time investment by activity—medium term

	Light treatment (LT)				Full treatment (FT)				LT=FT	Obs
	β	Unadj.	WILD	Rom. Wolf	β	Unadj.	WILD	Rom. Wolf	Unadj.	
Play	0.3597	0.0016	0.0135	0.0006	0.4112	0.0010	0.0126	0.0000	0.3545	1102
Sing	0.2200	0.0015	0.0087	0.0014	0.2601	0.0001	0.0100	0.0000	0.3524	1103
Tell stories	0.2157	0.0249	0.0486	0.0388	0.2661	0.0032	0.0128	0.0044	0.5692	1103
Visit relatives	0.2527	0.0032	0.0108	0.0032	0.2780	0.0104	0.0440	0.0096	0.7753	1103
Teach something new	0.3326	0.0125	0.0127	0.0218	0.3087	0.0516	0.0819	0.0562	0.8754	1103
Count	0.2082	0.0384	0.0801	0.0616	0.3596	0.0005	0.0100	0.0000	0.1090	1103
Hug	0.2946	0.0022	0.0115	0.0018	0.2666	0.0149	0.0147	0.0140	0.6706	1103
Respond verbally	0.0730	0.1667	0.1978	0.1862	-0.0608	0.1867	0.2261	0.2108	0.0177	1102
Praise	0.2450	0.0000	0.0062	0.0000	0.1579	0.0046	0.0319	0.0082	0.0258	1102
Positive discipline	0.1200	0.0348	0.0494	0.0616	0.1532	0.0130	0.0132	0.0128	0.3702	1103
Do not shout	0.0601	0.5303	0.5600	0.5181	0.0901	0.0899	0.1302	0.1358	0.7616	1103
Do not hit	0.1473	0.0910	0.1222	0.1184	0.0668	0.3440	0.4795	0.2913	0.4340	1103
Read	0.5519	0.0000	0.0070	0.0000	0.7468	0.0002	0.0114	0.0000	0.1602	1103

Notes: the table presents estimates of treatment effects on mother time investment in each activity. The sample includes caregivers surveyed in the follow-up survey (2018). All estimates show results from separate OLS regressions based on equation (1). The dependent variables are the single components of the mother time investment index defined in Section 3 and in Appendix A.3. All regressions include control variables as defined in Table 1 and sampling weights. Columns 1 and 5 report respectively the light treatment and full treatment coefficients for each dependent variable. Columns 2 and 6 report the unadjusted two-tailed p-values. Columns 3 and 7 show the WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution (Cameron et al. 2008; Roodman et al. 2019). Columns 4 and 8 show the two tailed p-values from a 5,000 replications Romano Wolf step-down procedure (Romano and Wolf 2005; Clarke et al. 2019). Column 9 reports p-values of a t-test for the null hypothesis of no difference between the light treatment and full treatment groups. Column 10 shows the number of observations.

Source: authors' calculations based on data from the follow-up survey (2018).

Table B11: Mother's attitudes towards gender roles

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Control group - base									
Light treatment (LT)	0.011	-0.070	0.069*	0.022	0.214**	0.194**	0.227*	0.154	0.119*
	(0.029)	(0.071)	(0.034)	(0.070)	(0.073)	(0.082)	(0.112)	(0.089)	(0.060)
Full treatment (FT)	0.063*	-0.022	0.225***	0.275***	0.172	0.185	0.338**	0.083	0.194**
	(0.030)	(0.058)	(0.054)	(0.065)	(0.121)	(0.146)	(0.116)	(0.067)	(0.074)
WILD p-values LT	0.707	0.478	0.109	0.777	0.031	0.068	0.120	0.182	0.111
WILD p-values FT	0.107	0.768	0.012	0.021	0.265	0.341	0.067	0.312	0.038
Romano-Wolf p-values LT	0.923	0.584	0.107	0.923	0.037	0.079	0.135	0.199	0.141
Romano-Wolf p-values FT	0.098	0.679	0.002	0.001	0.361	0.393	0.010	0.393	0.024
Observations	1105	1105	1105	1105	1105	1105	1105	1105	1105
R^2	0.015	0.010	0.022	0.029	0.040	0.051	0.037	0.035	0.026

Notes: the table presents the treatment effects on mother attitude. The sample includes caregivers surveyed in the follow-up survey (2018). All estimates show results from OLS regressions based on equation (1). All regressions include control variables as defined in Table 1 and sampling weights. The dependent variables in columns 1 to 9 include standardized z-scores of the 9 HOME-SF statements included in the attitudes towards gender roles section of the follow-up survey. Each statement is self-reported by the mother of the child. Each component is calculated by subtracting the control group mean and dividing by the control group standard deviation in each survey wave. Each column includes answers to each component of the attitudes towards gender roles section. Answers to negative statements are reverse coded such that a positive coefficient should be interpreted as a mother having more positive attitudes towards gender roles (less conservative). The statements of each component are the followings: fathers are less capable of looking after young children (column 1); if a man looks after the children people will laugh at him (column 2); the mother should be responsible for the home and the father should always make the important decisions (column 3); it is the main job of a woman to take care of children and household chores (column 4); fathers should have a closer relationship with their sons and mothers with daughters (column 5); women must seek partner's permission before going to visit the health center (column 6); girls should be at home most of the time (column 7); if a husband beats his wife it is because she deserves it as there needs to be order in the household (column 8); a daughter will be taken care of by another family when she is older so families need to ensure sons have all they need for a successful future and family (column 9). A full description of the construction of the outcomes is in Section 3 and in Appendix A.3. Light treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the light treatment group. Full treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the full treatment group.
*p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution are reported below the estimates (Cameron et al. 2008; Roodman et al. 2019). Two tailed p-values from a 5,000 replications Romano Wolf step-down procedure (Romano and Wolf 2005; Clarke et al. 2019) are shown below the estimates. P-values of a t-test for the null hypothesis of no difference between the light treatment and full treatment groups are also reported, number of observations and the R squared are presented at the bottom of the table.

Source: authors' calculations based on data from the follow-up survey (2018).

Table B12: Heterogeneous treatment effect—medium term

	(1) LT	(2) FT	(3) LTXBaseline characteristic	(4) FTXBaseline characteristic	(5) Baseline characteristic	(6) t-test (3) = (4)	(7) Obs
Panel A: *Heterogeneity on child development*							
Child is a girl	0.000	0.121	0.096	0.104*	0.010	0.902	1256
Child age $\geq$ 12 months	0.118	0.228**	-0.098	-0.075	0.051	0.849	1256
Asset index $\geq$ median	-0.015	0.142	0.086	0.045	-0.160	0.503	1256
Mother completed primary school	0.031	0.172	0.051	0.018	0.035	0.758	1256
Father completed primary school	0.031	0.185*	0.048	-0.015	0.107	0.447	1256
Panel B: *Heterogeneity on mother time investment*							
Child is a girl	0.172**	0.189**	0.069	0.038	-0.033	0.493	1103
Child age $\geq$ 12 months	0.223***	0.236**	-0.021	-0.037	0.005	0.674	1103
Asset index $\geq$ median	0.116	0.115	0.114	0.114	-0.136*	0.988	1103
Mother completed primary school	0.255***	0.241***	-0.107	-0.083	0.124	0.731	1103
Father completed primary school	0.230***	0.229***	-0.049	-0.047	0.073**	0.985	1103
Panel C: *Heterogeneity on mother influence*							
Child is a girl	0.108	0.131	-0.076	0.012	-0.009	0.417	1105
Child age $\geq$ 12 months	0.171	0.187	-0.155	-0.079	0.100	0.406	1105
Asset index $\geq$ median	0.007	0.106	0.076	0.034	-0.102	0.772	1105
Mother completed primary school	0.052	0.124	0.037	0.032	-0.050	0.955	1105
Father completed primary school	0.064	0.113	0.013	0.058	-0.080	0.832	1105

Notes: the table presents the heterogeneous treatment effects on child development and parenting outcomes. The sample includes children and caregivers surveyed in the follow-up survey (2018). Each row shows results from an OLS regression based on equation (1), and include an interaction term between each treatment dummy (LT and FT) and the following variables: child gender, child age, the asset index, the caregiver's education level defined as a binary variable equal to 1 if the caregiver (both mother and father) has at least primary education and 0 otherwise. In addition, all regressions control for the dependent variable at baseline, the total number of children in the household, the caregiver's age, and the caregiver's marital status. The definition of all control variables is provided in Appendix A.4. All regressions include sampling weights. The dependent variables in Panels A, B, and C are the child development aggregate index, the mother time investment aggregate index, and the mother influence aggregate index, respectively. A detailed definition of the dependent variables is provided in Section 3 and in Appendix A.3. Light treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the light treatment group. Full treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the full treatment group.
*p <10%, **p <5%, ***p <1% reported as WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution (Cameron et al. 2008; Roodman et al. 2019). Columns 1 and 2 reports the LT and FT effects. Column 3 reports coefficients on the interaction between the LT dummy and the baseline variable. Column 4 reports coefficients on the interaction between the FT dummy and the baseline variable. Column 5 reports coefficients on the baseline variable. Column 6 reports (WILD) p-values from the t-test on the difference between the coefficients in columns 3 and 4. Column 7 reports the number of observations.

Source: authors' calculations based on data from the follow-up survey (2018).

Table B13: Heterogeneity on child development at baseline—medium term

	(1) Child development	(2) Mother time investment	(3) Mother self-efficacy
Base group: Control group			
Light treatment (LT)	0.024	0.225***	0.185
	(0.137)	(0.055)	(0.117)
Full treatment (FT)	0.061	0.268***	0.257
	(0.139)	(0.072)	(0.158)
Light treatment (LT) × Child development baseline: 25-50th	-0.048	-0.012	-0.115
	(0.091)	(0.086)	(0.188)
Light treatment (LT) × Child development baseline: 50-75th	0.066	-0.065	-0.303*
	(0.050)	(0.107)	(0.139)
Light treatment (LT) × Child development baseline: above 75th	0.164	0.004	-0.073
	(0.114)	(0.069)	(0.139)
Full treatment (FT) × Child development baseline: 25-50th	0.169	-0.082	-0.082
	(0.112)	(0.078)	(0.235)
Full treatment (FT) × Child development baseline: 50-75th	0.253**	-0.104	-0.319*
	(0.082)	(0.106)	(0.158)
Full treatment (FT) × Child development baseline: above 75th	0.114	-0.051	-0.120
	(0.143)	(0.105)	(0.176)
Child development at baseline: 25-50th	0.100**	-0.022	0.036
	(0.041)	(0.075)	(0.163)
Child development at baseline: 50-75th	0.133***	0.043	0.211**
	(0.019)	(0.099)	(0.082)
Child development at baseline: above 75th	0.161***	0.034	0.125
	(0.028)	(0.066)	(0.094)
WILD p-values LT X Child development at baseline 25-50th	0.660	0.822	0.633
WILD p-values LT X Child development at baseline 50-75th	0.344	0.601	0.112
WILD p-values LT X Child development at baseline above 75th	0.305	0.967	0.664
WILD p-values FT X Child development at baseline 25-50th	0.311	0.471	0.776
WILD p-values FT X Child development at baseline 50-75th	0.056	0.410	0.074
WILD p-values FT X Child development at baseline above 75th	0.511	0.661	0.572
Observations	1256	1102	1104
R^2	0.068	0.085	0.023

Notes: the table presents the heterogeneous treatment effects on child development and parenting outcomes. The sample includes children and caregivers surveyed in the follow-up survey (2018). Each column shows results from an OLS regression based on equation (1), and also includes an interaction term between each treatment dummy (LT and FT) and the child development aggregate index at baseline. This is calculated by taking the average of the five ASQ z-scores components at baseline. The variable is then defined as follows: equal to 0 if the child development index at baseline is lower than the top 25th percentile, 1 if the child development index at baseline is between the 25th and 50th percentile, 2 if the child development index at baseline is between the 50th and 75th percentile; and 3 if the child development index at baseline is above the 75th percentile. All regressions control for the dependent variable at baseline, child gender, child age, the total number of children in the household, the caregiver's age, the caregiver's marital status. The definition of all control variables is provided in Appendix A.4. All regressions include sampling weights. The dependent variable in column 1 is the child development aggregate index, in column 2 is the mother time investment aggregate index, and in column 3 is the mother influence aggregate index. A full description of the construction of the outcomes is in Section 3 and in Appendix A.3. Light treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the light treatment group. Full treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the full treatment group.
*p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution are reported below the estimates (Cameron et al. 2008; Roodman et al. 2019).

Source: authors' calculations based on data from the follow-up survey (2018).

Table B14: Crowding-out of parents investments

Panel A: Short term	Child development	Mother time investment	Mother influence
Control group - base			
Light treatment (LT)	0.275*	0.421***	0.449***
	(0.131)	(0.046)	(0.072)
Full treatment (FT)	0.378**	0.612***	0.574***
	(0.142)	(0.036)	(0.075)
LT X Chilren below 3 years in HH	0.017	0.053	-0.030
	(0.054)	(0.030)	(0.050)
FT X Chilren below 3 years in HH	0.003	0.011	0.020
	(0.041)	(0.013)	(0.022)
Number of children below 3 in the HH	-0.018	-0.027	-0.035
	(0.026)	(0.014)	(0.023)
WILD p-values LT interaction	0.816	0.183	0.653
WILD p-values FT interaction	0.962	0.431	0.392
Observations	1428	1299	1300
R^2	0.168	0.267	0.106
Panel B: Medium term	Child development	Mother time investment	Mother self-efficacy
Control group - base			
Light treatment (LT)	0.111	0.190***	-0.055
	(0.105)	(0.022)	(0.063)
Full treatment (FT)	0.298**	0.209**	0.069
	(0.107)	(0.067)	(0.082)
LT X Chilren below 3 years in HH	-0.053	0.020	0.101*
	(0.036)	(0.023)	(0.046)
FT X Chilren below 3 years in HH	-0.095**	0.002	0.050
	(0.032)	(0.033)	(0.040)
Number of children below 3 in the HH	0.039	-0.019	-0.034
	(0.031)	(0.018)	(0.020)
WILD p-values LT interaction	0.305	0.558	0.082
WILD p-values FT interaction	0.032	0.940	0.273
Observations	1256	1103	1105
R^2	0.082	0.079	0.020

Notes: the table presents results from a test of crowding-out of parents' investments. The sample includes children and caregivers surveyed in the end-line survey (Panel A) and in the follow-up survey (Panel B). All estimates show results from OLS regressions based on equation (1). All regressions include control variables as defined in Table 1 and sampling weights. All control variables are described in Appendix A.4. The specification includes also an interaction term between the light treatment and the full treatment variables and a continuous variable defined as the number of children below 3 years old living in the house. The dependent variables are the child development index (column 1, Panel A and B), the mother time investment index (column 2, Panel A and B), the mother influence index (column 3, Panel A) and the mother self-efficacy index (column 3, Panel B). A full description of the construction of the outcomes is in Section 3 and in Appendix A.3. Light treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the light treatment group. Full treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the full treatment group.
*p <10%, **p <5%, ***p <1%. Robust standard errors in parenthesis are clustered at the sector level. WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution are reported below the estimates cameron2008bootstrap, roodman2019fast. The number of observations and the R squared are reported at the bottom of the table.

Source: authors' calculations based on data from the end-line (2016) and follow-up (2018) surveys.

Table B15: Mechanisms

	Communication		Gross motor		Fine motor		Problem solving		Personal social		Child development index	
Mother time investment	0.346***	0.287***	0.339***	0.278***	0.248***	0.146**	0.339***	0.270***	0.444***	0.395***	0.327***	0.263***
	(0.059)	(0.060)	(0.072)	(0.080)	(0.050)	(0.055)	(0.066)	(0.072)	(0.038)	(0.036)	(0.040)	(0.044)
Mother influence		0.080*		0.083		0.137**		0.093**		0.065*		0.087**
		(0.037)		(0.045)		(0.048)		(0.036)		(0.034)		(0.033)
WILD pvalues time investment	0.001	0.006	0.002	0.014	0.001	0.047	0.001	0.003	0.000	0.000	0.001	0.001
WILD pvalues influence		0.048		0.086		0.016		0.037		0.104		0.009
Observations	1299	1299	1299	1299	1299	1299	1299	1299	1299	1299	1299	1299
R^2	0.199	0.202	0.109	0.113	0.084	0.093	0.075	0.079	0.123	0.125	0.180	0.188

Notes: the table presents results from an OLS regression that correlates mother time investment and mother influence with child development outcomes. All regressions include control variables as defined in Table 1 and sampling weights. All control variables are described in Appendix A.4. The dependent variables in columns 1 to 10 include standardized z-scores of the five ASQ components calculated by subtracting the control group mean and dividing by the control group standard deviation in each survey wave. The dependent variable in columns 11 and 12 is the child development index-calculated by taking the average of the five ASQ z-scores. All regressions include also as main regressors the mother time investment aggregate index. Columns 2, 4, 6, 8, 10, and 12 include also the mother influence aggregate index. A full description of the construction of the outcomes is in Section 3 and in Appendix A.3.
*p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution are reported below the estimates (Cameron et al. 2008; Roodman et al. 2019). Observations and the R squared are presented at the bottom of the table.

Source: authors' calculations based on data from the end-line survey (2016).

C Validity and robustness checks

C.1 Attrition

We tested for differential attrition in the end-line and follow-up surveys. At end-line, the attrition rate of caregivers was ten per cent. At follow-up, the attrition rate was 18 per cent. These attrition rates after almost three years from the intervention are consistent with rates found in similar early child development programs in sub-Saharan Africa (Britto et al. 2017). Most of the families and children not found at end-line or follow-up moved away. This can generate two potential biases. First, selection bias may occur if across intervention groups those families who moved away have some common observable and unobservable characteristics related to child development or parenting practices. Second, results may be biased if there is a higher likelihood of dropping out from the sample in specific intervention arms in relation to the other arms. To address these potential biases, we tested for whether attrition rates are balanced by treatment status.

The dependent variable in Table C1 is defined as 1 if the child dropped out at end-line (columns 1 and 2) or follow-up (columns 3 and 4). We included in the specification light and full treatment dummies, child and household characteristics, and interacted each characteristic with the treatment dummies. The results in Table C1 indicate that overall there does not seem to be differential attrition between treatment groups, with the exception of a lower probability of attrition at end-line in the full treatment group (FT), as shown by the coefficient on FT in column 1. However, this effect disappears when we include child and household characteristics in the model (column 2). The joint F-test of interactions also suggests that there are no differences in the characteristics of children who drop out of the sample at follow-up. The only difference present in the full treatment at end-line, disappears at follow-up. These results reassure us that selection bias caused by attrition is unlikely to affect substantially the validity of the results presented in the paper.

C.2 Self-reporting and social desirability bias

The variables we use to measure parental time investment and some dimensions of child development are self-reported. This could introduce social desirability bias in the answers given by the respondent, which could potentially raise concerns about measurement error if the respondent provides biased answers because they joined the treatment. In order to test for social desirability bias in reporting, we looked at whether the effects of the program on parental time investments differed depending on whether the response was self-reported or reported by their partner. In Panel A of Table C2, we compared father's time investment self-reported by the father (column 1) to father's time investments as reported by the mother (column 2). We also estimated levels of maternal time investments self-reported by the mother (column 3) compared to maternal time investments reported by the father (column 4). The main assumption underlying these comparisons is that respondents may be more prone to social desirability biases when answering questions about their own behaviour. We find that the effect of the programme is very similar across both types of answers. We acknowledge that this test may only partially address potential social desirability biases in self-reported answers, but the results of this analysis suggest that the magnitude of any potential bias is likely to be small with negligible effects on the validity of our main results.

We also tested for potential social desirability bias in self-reported child development outcomes. Most of the tasks that generated information about child development outcomes were reported by the enumerator while observing the child performing a certain activity/task. In one domain of the ASQ questionnaire—personal social skills—the answers were partly self-reported by the caregiver, and partly observed by the enumerator. This particular feature of the ASQ questionnaire allows us to test for potential social desirability by comparing both answers. Panel B of Table C2 shows the treatment effect on the ASQ personal social skills indicator. Column 1 reports the result for self-reported questions, and column 2

reports results for answers observed by the enumerator. The results are largely similar in both columns, which mitigates possible concerns about measurement error caused by social desirability bias.

C.3 Testing imbalance

We performed several checks to account for the fact that some characteristics and outcome variables were not balanced at baseline, as discussed in Section 3.3.

First, we checked whether the source of imbalance that we observe at baseline mostly in the LT group is driven by a specific age group, gender, or ASQ activity. We find that the imbalance is mostly attributable to children who were between ten and 15 months old, who were administered the 12 months ASQ questionnaire. As a robustness check, we excluded from the sample these children, and results remain unchanged (Table C3).

Second, following Baranov et al. (2020) we estimated a model in which we included the full set of control variables and the ASQ aggregate index at baseline, demeaned and interacted with the treatment indicator. This accounts for the fact that some variables were not balanced at baseline. The interaction with the treatment variable allows for differing impacts of these characteristics on outcomes. Results shown in the Table C4 are similar to the main results discussed in Section 5.

Third, we estimated a model using inverse probability weighting (Cattaneo 2010). As shown in Table B1, we observe that the ASQ scores, the proportion of children who are female, the total number of children in the household, the respondent age, and the mother's education are not balanced between the control group and the light treatment, and in some cases also in the full treatment group. These characteristics at baseline may be dependent on the treatment assignment. We estimated a model that gives less weight to the observations that in the LT or FT group show higher ASQ scores, have larger proportion of female children, a larger number of children, with older and more educated respondents and, at the same time, attributes more weight to the observations in the control group with the opposite characteristics (e.g. lower proportion of female children, and so on). We estimated the probability of being in the LT or FT groups (using a multinomial logit model) on the variables that showed an imbalance at baseline, and used the resulting predictions to weight the observations. Results in Table C5 are largely consistent with our main results, although the size of the coefficient is slightly smaller.

C.4 Multiple hypothesis testing

In the main model specification, we estimate the impact of First Steps on single dimensions of the child development and parental outcomes aggregated at the mean index. We also show estimates for each dimension accounting for Family Wise Error Rate (FWER) at the bottom of the main tables using the Romano and Wolf (2005) correction in order to account for multiple hypothesis testing.

The procedure starts by placing each outcome of interest in a family of related outcomes. In our case, one group of outcomes includes child development outcomes, whilst another group includes parental time investment outcomes. A third group includes parental influence outcomes. The procedure then calculates a t-statistic of our main hypothesis: that the treatment has an effect on the outcome. This is calculated for each outcome, and the calculated t-statistics are subsequently ranked from the largest to the smallest within each family. The largest observed t-statistic is then compared with the distribution of the maximal bootstrapped t-statistics. The null distribution for the re-sampling is based on the standard error of each parameter estimated in the original model, in line with Romano and Wolf (2005)—see algorithm 4.2. The distribution of bootstrap t-statistics results from 5,000 replications of the t-statistic of the main hypothesis. The distribution is composed by the highest t-statistics in each of the 5,000 replications. The reported Romano Wolf p-value is the probability of observing the original t-statistic larger than the bootstrap distribution of t-statistics. If this probability is high ($p \geq 0.1$) we fail to reject

the joint null hypothesis that the treatment has no impact on any outcome in the family of hypotheses being tested. If the probability is below p=0.1, we reject this joint null hypothesis, remove the most significant hypothesis, and test the subset of hypotheses that remain for joint significance (the ones with a lower original t-statistics). This process of dropping the most significant hypothesis continues until the resulting subset of hypotheses fails to be rejected, or only one hypothesis remains. This method is superior to the Bonferroni adjustment method as it accounts for interdependence across outcomes.

C.5 Weighted mean index

Our main estimates reported in Tables 1 to 5 show results obtained using the mean index of child development and on parenting outcomes constructed as unweighted mean indices. This implies that each dimension or activity is given the same weight in the index. Following Anderson (2008), we also estimated our models using an inverse co-variance weighted index. Considering a group of different outcomes, this procedure assigns less weights to the outcomes within the group which are highly correlated, while it rewards new information by giving a higher weight to outcomes that are less correlated within the same group. The weight for each outcome is the sum of the inverted cov-ariance matrix that includes all outcomes in the group considered. Results in Table C6 are largely consistent with the main estimates. We observe, however, that the effect sizes on all outcomes and treatment group are larger than those using unweighted indexes. Although the inverse co-variance weighted index has some merit, most of the existing literature has shown a preference towards using the unweighted mean index because the statistical procedure used to assign the weights is not yet conclusive (Glennerster and Takavarasha 2013). Therefore, our main estimates use the unweighted mean index (Kling et al. 2007) and we present the weighted estimates here for completeness and comparison.

C.6 Mediation analysis

In this section, we discuss the implementation of a linear mediation analysis, and estimate a linear equation model to decompose the treatment effect. Following Heckman et al. (2013) and Kim et al. (2018), we estimated the following linear model:

$$E[Y_1 - Y_0] = a^j E[P_1^J - P_0^J] + \tau_1 - \tau_0 + (b_1 - b_0)X_{k'} \tag{2}$$

where $E[Y_1 - Y_0]$ is the estimated change in the outcome; $E[P_1^J - P_0^J]$ is the contribution of the observed parental inputs (maternal time investment and maternal influence). $\tau_1 - \tau_0 + (b_1 - b_0)X_{k'}$ is the contribution of the unmeasured parental input.

We first tested whether there is any unobserved treatment effect on parental inputs. We regressed child development outcomes on the interaction between the treatment status (T=0 or T=1) and maternal inputs or the baseline covariates. Table C7 shows that the coefficients of the interaction terms between treatment and control groups are largely similar. These results suggest that there is no economically and statistically relevant unobserved treatment effect on observed parental inputs and covariates. Table C8 shows the same test for the analysis using the follow-up data. The results show statistically significant differences in some dimensions. Thus, we present the results of the mediation analysis at 33 months, but acknowledge that these results need to be interpreted with caution.

In order to obtain estimates for $E[P_1^J - P_0^J]$ for each parental input we observe, we regressed each parental input (i.e. maternal time investment and maternal influence) on the treatment dummy, while controlling for baseline covariates. We then regressed all child development outcomes on the treatment and parental inputs, adding the usual baseline covariates. From these estimates we obtain the a^j coefficients which measure the direct impact of each parental input on child development outcomes. The coefficient of the treatment term measures all other unobserved factors that influence child development (i.e. other than our observed parental inputs measures)

The treatment effect is then decomposed into experimentally induced changes in measured and unmeasured inputs. This decomposition allows us to extract the individual effect of observed parental inputs (maternal time investment and maternal influence) and other unobserved factors.

Table C1: Attrition

	End-line		Follow-up	
Dependent: Obs. lost at end-line and follow-up				
Control group				
Light treatment (LT)	-0.015	-0.147	-0.023	-0.065
	(0.018)	(0.209)	(0.023)	(0.180)
Full treatment (FT)	-0.036**	-0.210	0.022	0.155
	(0.015)	(0.141)	(0.018)	(0.147)
Child characteristics				
Child is a female		-0.024		-0.034
		(0.025)		(0.025)
Child age in months		0.000		-0.002
		(0.001)		(0.003)
Total number of children		0.002		-0.005**
		(0.008)		(0.002)
HH characteristics				
Caregiver age		-0.004		-0.002*
		(0.003)		(0.001)
Mother has at least primary education		-0.040		0.004***
		(0.035)		(0.001)
Father has at least primary education		0.038		0.097**
		(0.028)		(0.042)
Respondent is married or cohabitating		0.010		-0.021
		(0.047)		(0.068)
HH asset index		-0.017		-0.011
		(0.011)		(0.017)
Interactions: LT X Child characteristics				
LT X Child is a girl		0.036		0.028
		(0.029)		(0.028)
LT X Child age in months		0.002		0.008**
		(0.003)		(0.003)
LT X Children in the HH		-0.004		-0.007
		(0.017)		(0.019)
Interactions: LT X HH characteristics				
LT X Respondent age		0.004		0.002
		(0.004)		(0.002)
LT X Mother has at least primary education		0.045		0.026
		(0.049)		(0.023)
LT X Father has at least primary education		-0.036		-0.081
		(0.032)		(0.054)
LT X Respondent is married or cohabitating		-0.030		-0.029
		(0.063)		(0.097)
LT X Asset index		-0.003		-0.014
		(0.015)		(0.022)
Interactions: FT X Child characteristics				
FT X Child is a girl		0.030		0.080*
		(0.033)		(0.040)
FT X Child age in months		0.000		-0.000
		(0.001)		(0.003)
FT X Children in the HH		-0.004		0.008
		(0.009)		(0.014)
Interactions: FT X HH characteristics				
FT X Respondent age		0.001		-0.003
		(0.003)		(0.002)
FT X Mother has at least primary education		0.050		-0.016
		(0.044)		(0.014)

Table C1: Attrition

	End-line		Follow-up	
FT X Father has at least primary education		0.016		-0.040
		(0.056)		(0.071)
FT X Respondent is married or cohabitating		0.022		-0.016
		(0.070)		(0.074)
FT X Asset index		0.009		-0.005
		(0.013)		(0.020)
Joint F-test of Interactions (p-value)				
With LT		0.185		0.434
With FT		0.028		0.389
Observations	1614	1583	1614	1583

Notes: the table presents the test on differential attrition at end-line and follow-up. All estimates show results from OLS regressions. In columns 1 and 2, the dependent variable (child lost at end-line) is defined as 1 if the observation at end-line is missing, and 0 otherwise. In columns 3 and 4, the dependent variable (child lost at follow-up) is defined as a dummy variable equal to 1 if the observation at follow up is missing, and 0 otherwise. Regressions in columns 1 and 3 include as independent variables the light treatment dummy defined as equal to 1 if the respondent (the caregiver) participated in the light treatment group; the full treatment dummy defined as equal to 1 if the respondent (the caregiver) participated in the full treatment group. Regressions in columns 2 and 4 include also the following controls: child gender, child age, the total number of children in the household, the caregiver's age, the caregiver's education level, the caregiver's marital status, and the asset index. Regressions in columns 2 and 4 include also the interaction terms between the light and full treatment dummies and each of the control variables. All regressions include sampling weights.

*p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. Joint F-tests of the interaction between each treatment arm and the control variables at baseline are reported at the bottom of the tables. Observations are reported at the bottom of the table.

Source: authors' calculations based on data from the end-line (2016) and follow-up (2018) surveys.

Table C2: Social desirability bias

Panel A: Parents investment	(1) Father time investment (self reported)	(2) Father time investment (reported by mother)	Mother time investment (self reported)	Mother time investment (reported by father)
Treatment (LT and FT)	0.339**	0.399***	0.546***	0.376***
	(0.140)	(0.035)	(0.040)	(0.090)
WILD p-values treatment	0.015	0.002	0.000	0.007
Observations	87	1299	1299	87
R^2	0.260	0.210	0.255	0.278

Panel B: ASQ personal social skills	(1) Self reported	(2) Enumerator reported
Treatment (LT and FT)	0.255***	0.286***
	(0.050)	(0.068)
WILD p-values treatment	0.005	0.016
Observations	919	919
R^2	0.523	0.193

Notes: the table presents results from a regression that estimates the treatment effect (combined LT and FT effect) on parenting and child development outcomes to test for social desirability bias. All estimates show results from OLS regressions. All regressions control for baseline values of the outcome of interest, child gender, child age, the total number of children in the household, the caregiver's age, the caregiver's education level, the caregiver's marital status, and the asset index. All controls are described in detail in Appendix A.4. All regressions include sampling weights. The dependent variables in Panel A are: father time investment aggregate index self-reported by the father (column 1); father time investment aggregate index reported by the mother (column 2); mother time investment aggregate index self-reported by the mother (column 3); mother time investment aggregate index reported by the father (column 4). The dependent variables in Panel B are: ASQ personal social skills component self-reported by the caregiver (column 1); ASQ personal social skills component reported by the enumerator (2). The two dependent variables are calculated by taking the average of the ASQ personal social skills z-scores. A full description of the construction of these outcomes is in Appendix C.2. Treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in either the light treatment group or the full treatment group.
*p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution are reported below the estimates (Cameron et al. 2008; Roodman et al. 2019). Observations and the R squared are presented at the bottom of the table.

Source: authors' calculations based on data from the end-line survey (2016).

Table C3: Child development—testing imbalance

Panel A: Short term	(1) Child development (full sample)	(2) Child development (excluding ASQ 12 months)
Control group - base		
Light treatment (LT)	0.292**	0.264**
	(0.087)	(0.089)
Full treatment (FT)	0.379***	0.368***
	(0.094)	(0.092)
WILD p-values LT	0.022	0.027
WILD p-values FT	0.015	0.015
t-test LT = FT		
p-value	0.041	0.002
Observations	1428	920
R^2	0.168	0.162

Panel B: Medium term	(1) Child development (full sample)	(2) Child development (excluding ASQ 12 months)
Control group - base		
Light treatment (LT)	0.051	0.063
	(0.077)	(0.079)
Full treatment (FT)	0.181**	0.181**
	(0.075)	(0.058)
WILD p-values LT	0.590	0.538
WILD p-values FT	0.092	0.038
t-test LT = FT		
p-value	0.069	0.040
Observations	1256	809
R^2	0.078	0.081

Notes: the table presents the treatment effects on child development outcomes. Column 1 presents the light and full treatment effects calculated on the full sample. Column 2 presents the light and full treatment effects on a sample where we removed children aged between 10 and 15 months old at baseline who have been administered the 12 months ASQ questionnaire. All estimates show results from OLS regressions. All regressions include control variables as defined in Table 1 and sampling weights. The dependent variable in columns 1 and 2 is the child development aggregate index. A full description of the construction of the outcomes is in Section 3 and in Appendix A.3. Light treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the light treatment group. Full treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the full treatment group.
*p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution are reported below the estimates (Cameron et al. 2008; Roodman et al. 2019). P-values of a t-test for the null hypothesis of no difference between the light treatment and full treatment groups are also reported, number of observations and the R squared are presented at the bottom of the table.
Source: authors' calculations based on data from the end-line survey (2016) for Panel A and follow-up survey (2018) for Panel B.

Table C4: Testing for imbalance of control variables

Panel A: Short term	(1) Child development	(2) Mother time investment	(3) Mother influence
Control group - base			
Light treatment (LT)	0.294**	0.456***	0.410***
	(0.088)	(0.017)	(0.050)
Full treatment (FT)	0.365***	0.604***	0.577***
	(0.096)	(0.034)	(0.056)
WILD p-values LT	0.019	0.002	0.009
WILD p-values FT	0.013	0.004	0.008
t-test LT = FT			
p-value	0.102	0.002	0.011
Observations	1428	1298	1299
R^2	0.174	0.294	0.124

Panel B: Medium term	(1) Child development	(2) Mother time investment	(3) Mother self efficacy
Control group - base			
Light treatment (LT)	0.047	0.197***	0.035
	(0.080)	(0.007)	(0.056)
Full treatment (FT)	0.189**	0.195***	0.121**
	(0.079)	(0.034)	(0.050)
WILD p-values LT	0.659	0.002	0.660
WILD p-values FT	0.097	0.012	0.063
t-test LT = FT			
p-value	0.054	0.959	0.068
Observations	1256	1102	1104
R^2	0.088	0.099	0.031

Notes: the table presents the treatment effects on child development and parenting outcomes, controlling for imbalance at baseline. The specification includes the full set of baseline characteristics demeaned and interacted with the treatment variables. These are the baseline values of the dependent variables, child gender, child age, the total number of children in the household, the caregiver's age, the caregiver's education level, the caregiver's marital status, and the asset index. This accounts for the fact that the sample was not balanced along all observable baseline characteristics, and the interaction with treatment allows for differing impacts of these characteristics on outcomes. All estimates show results from OLS regressions. All regressions include sampling weights. The dependent variables are the child development index (column 1, Panel A and B), the mother time investment index (column 2, Panel A and B), the mother influence index (column 3, Panel A), and the mother self-efficacy index (column 3, Panel B). A full description of the construction of the outcomes is in Section 3 and in Appendix A.3. Light treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the light treatment group. Full treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the full treatment group. *p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution are reported below the estimates (Cameron et al. 2008; Roodman et al. 2019). P-values of a t-test for the null hypothesis of no difference between the light treatment and full treatment groups are also reported, number of observations and the R squared are presented at the bottom of the table.

Source: authors' calculations based on data from the end-line survey (2016) for Panel A and follow-up survey (2018) for Panel B.

Table C5: Inverse probability weighting

Panel A: Short term	Child development	Mother time investment	Mother influence
	(1)	(2)	(3)
Light treatment (with IPW weights)	0.290***	0.470***	0.414***
	(0.044)	(0.034)	(0.055)
Full treatment (with IPW weights)	0.349***	0.620***	0.567***
	(0.045)	(0.033)	(0.053)
Observations	1450	1319	1319
Panel B: Medium term	Child development	Mother time investment	Mother self-efficacy
	(1)	(2)	(3)
Light treatment (with IPW weights)	0.062	0.209***	0.045
	(0.047)	(0.030)	(0.054)
Full treatment (with IPW weights)	0.181***	0.204***	0.133**
	(0.046)	(0.032)	(0.057)
Observations	1275	1120	1120

Notes: the table presents the treatment effects on child development and parenting outcomes accounting for the baseline imbalance. We estimate the probability of being in the LT or FT groups (using a multinomial logit model) on the variables that showed an imbalance at baseline and use the resulting predictions to weight the observations. All estimates are OLS regressions based on specification (1). All regressions control for baseline values of the outcome of interest, child gender, child age, the total number of children in the household, the caregiver's age, the caregiver's education level defined as a binary variable equal to 1 if the caregiver has at least primary education and 0 otherwise, the caregiver's marital status defined as a binary variable equal to 1 if the caregiver is married or cohabitating and 0 otherwise, and the asset index which is equal to the first principal component of the variables floor materials of the house, roof materials of the house, main source of drinking water, and whether the house of the respondent is owned or rented, as described in detail in Appendix A.4. All regressions include sampling weights. In Panel A, the dependent variable in column 1 is the child development index calculated by taking the average of the five ASQ z-scores. The dependent variable in column 2 is the mother time investment index when self-reported by the mother of the child, calculated by taking the average of the three HOME-SF z-scores. The dependent variable in column 3 is the mother influence index when self-reported by the mother of the child, calculated by taking the average of the six HOME-SF z-scores. Panel B mirrors the same outcomes as Panel A, with the exception of the dependent variable in column 3, which is the mother self-efficacy index calculated by taking the average of the eight TOPSE z-scores. A full description of the construction of the outcomes is in Section 3 and in Appendix A.3. Light treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the light treatment group. Full treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the full treatment group.
*p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution are reported below the estimates (Cameron et al. 2008; Roodman et al. 2019). A t-test LT = FT is presented, with the statistical significance of the test expressed in p-value. Observations and the R squared are presented at the bottom of the table.

Source: authors' calculations based on data from the end-line survey (2016) for Panel A and follow-up survey (2018) for Panel B.

Table C6: Weighted mean index—short and medium term

Panel A: Short term	Child development	Mother time investment	Mother influence
	(1)	(2)	(3)
Light treatment (LT)	0.394**	0.742***	0.477***
	(0.122)	(0.063)	(0.059)
Full treatment (FT)	0.494***	1.022***	0.721***
	(0.132)	(0.080)	(0.074)
WILD p-values LT	0.024	0.004	0.008
WILD p-values FT	0.012	0.006	0.009
t-test LT = FT			
p-value	0.085	0.004	0.007
Observations	1428	1299	1300
R^2	0.176	0.184	0.100
Panel B: Medium term	Child development	Mother time investment	Mother self-efficacy
	(1)	(2)	(3)
Light treatment (LT)	0.080	0.490***	0.096
	(0.119)	(0.050)	(0.075)
Full treatment (FT)	0.284**	0.482***	0.086
	(0.119)	(0.053)	(0.070)
WILD p-values LT	0.597	0.006	0.313
WILD p-values FT	0.101	0.008	0.433
t-test LT = FT			
p-value	0.048	0.912	0.879
Observations	1256	1100	1105
R^2	0.077	0.068	0.014

Notes: the table presents the treatment effects on child development and parents outcomes using a weighted index. The sample includes children and caregivers surveyed in the end-line survey (Panel A) and in the follow-up survey (Panel B). All estimates show results from OLS regressions based on equation (1). All regressions include control variables as defined in Table 1 and sampling weights. All control variables are described in Appendix A.4. The dependent variables are the weighted child development index (column 1, Panel A and B), the mother time investment index (column 2, Panel A and B), the mother influence index (column 3, Panel A), and the mother self-efficacy index (column 3, Panel B). The weighting procedure assigns less weights to the components of an outcome which are highly correlated, while it rewards new information by giving a higher weight to components of an outcome that are less correlated within the same group. A full description of the construction of the outcomes is in Section 3 and in Appendix A.3. The weighting procedure is described in Appendix C.5. Light treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the light treatment group. Full treatment is a dummy variable equal to 1 if the respondent (the caregiver) participated in the full treatment group.
*p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. WILD cluster bootstrap with 99,999 replications and residuals drawn from Webb's 6-point distribution are reported below the estimates (Cameron et al. 2008; Roodman et al. 2019). P-values of a t-test for the null hypothesis of no difference between the light treatment and full treatment groups are also reported, number of observations and the R squared are presented at the bottom of the table.
Source: authors' calculations based on data from the end-line survey (2016) for Panel A and follow-up survey (2018) for Panel B.

Table C7: Mediation test—short term

	Communication (1)	Gross motor (2)	Fine motor (3)	Problem Solving (4)	Personal social (5)	Child development (6)
Control group × Mother time investment	0.206	0.166	0.015	0.210***	0.260***	0.154***
Treatment × Mother time investment	0.186**	0.318***	0.036	0.136**	0.314***	0.195***
Control group × Mother influence	0.072	0.113*	0.139**	0.052	0.075*	0.091*
Treatment × Mother influence	0.071	0.074	0.121	0.096*	0.045	0.077*
Control group × Child is a female	0.095	-0.068***	-0.107**	-0.121***	0.080	-0.017
Treatment × Child is a female	0.157**	-0.065	-0.049	0.014	0.032	0.013
Control group × Child age in months	0.072***	0.013***	-0.014	0.019*	-0.022*	0.015
Treatment × Child age in months	0.073***	0.011*	-0.008	0.015***	-0.009	0.017***
Control group × Total number of children	-0.015	-0.028	0.016	0.020	0.018	-0.001
Treatment × Total number of children	0.011	-0.010	0.010	-0.008	0.002	0.002
Control group × Caregiver age	0.006	0.009*	-0.005	-0.009	0.001	0.002
Treatment × Caregiver age	-0.001	0.012**	0.002	0.007	0.003	0.005
Control group × Mother highest education	0.136	0.192***	0.224	0.066	0.086	0.132
Treatment × Mother highest education	0.163**	0.035	0.099	0.078	0.060	0.084
Control group × Father highest education	0.048	-0.014	0.136	0.066	0.201***	0.072
Treatment × Father highest education	0.071*	0.023	0.119*	0.116	0.083	0.082
Control group × Marital status	0.121	0.126	0.063	0.214*	0.095	0.119*
Treatment × Marital status	0.268***	0.147**	0.151	0.109	0.243***	0.170***
Control group × HH asset index	-0.050***	-0.028	-0.018	-0.082***	-0.004	-0.039**
Treatment × HH asset index	-0.029*	-0.029*	-0.002	-0.013	-0.028*	-0.020
ASQ communication at baseline	0.123***					
ASQ gross motor at baseline		0.177***				
ASQ fine motor at baseline			0.155***			
ASQ problem solving at baseline				0.098***		
ASQ personal social at baseline					0.161***	
Child development index at baseline						0.215***
t-test mother time investment (p-value)	0.888	0.312	0.847	0.243	0.307	0.529
t-test mother influence (p-value)	0.997	0.633	0.835	0.509	0.589	0.822
Observations	1299	1299	1299	1299	1299	1299
R^2	0.212	0.115	0.107	0.098	0.138	0.202

Notes: the table tests whether there is unobserved treatment effect on the observed parental inputs (mother time investment and mother influence) as explained in Appendix C.6. All estimates are OLS regressions. The table reports coefficients from a specification that regresses the child development outcomes on the interaction between the treatment status (LT=0 and FT=0, or LT and FT=1) and maternal inputs or the baseline covariates. These are child gender, child age, the total number of children in the household, the caregiver's age, the caregiver's education level, the caregiver's marital status, and the asset index. These variables are described in Appendix A.4. Mother time investment index self-reported by the mother of the child is calculated by taking the average of the three HOME-SF z-scores. Mother influence self-reported by the mother of the child is calculated by taking the average of the six HOME-SF z-scores. All regressions include sampling weights. The dependent variables in columns 1 to 5 include standardized z-scores of the five ASQ dimensions calculated by subtracting the control group mean and dividing by the control group standard deviation in each survey wave. The dependent variable in column 6 is the child development index calculated by taking the average of the five ASQ z-scores. A full description of the construction of the outcomes is in Section 3 and in Appendix A.3. Treatment is a binary variable equal to 1 if the respondent (the caregiver) participated in either the light treatment group or the full treatment group. Control is a binary variable equal to 1 if the respondent (the caregiver) did not participate in the treatment.
*p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. P-values of a t-test for the null hypothesis of no difference between the interaction terms 'Control group X Mother Time Investment' and 'Treatment X Mother Time Investment', and 'Control group X Mother Influence' and 'Treatment X Mother Influence' are reported at the bottom of the table along with the observations and the R squared.

Source: authors' calculations based on data from the end-line survey (2016).

Table C8: Mediation test—medium term

	ASQ Communication (1)	ASQ Gross motor (2)	ASQ Fine motor (3)	ASQ Problem solving (4)	ASQ Personal Social (5)	Child development (6)
Control group × Mother time investment	0.331**	0.355*	0.261***	0.328***	0.394**	0.330***
Treatment × Mother time investment	0.422***	0.257***	0.312***	0.333***	0.362***	0.330***
Control group × Mother self-efficacy	-0.012	0.002	0.024	0.156***	0.210***	0.068***
Treatment × Mother self-efficacy	-0.028	-0.018	0.013	0.048	0.010	0.004
Control group × Child is a female	-0.078	0.035	-0.119***	0.053	0.048	-0.009
Treatment × Child is a female	0.035	0.055	0.049	0.219***	0.151**	0.095**
Control group × Child age in months	0.029**	0.022***	0.050***	-0.050***	0.005	0.012**
Treatment × Child age in months	0.019***	0.011**	0.043***	-0.063***	0.007	0.004
Control group × Total number of children	0.062	0.147*	0.033	0.041	0.074	0.069
Treatment × Total number of children	-0.037	0.004	0.008	0.012	0.020	0.003
Control group × Caregiver age	-0.007	-0.026	-0.004	0.003	-0.013	-0.009
Treatment × Caregiver age	-0.001	-0.003	-0.001	0.004	-0.005	-0.001
Control group × Mother highest education	0.049**	0.003	0.219***	0.060	-0.149**	0.028
Treatment × Mother highest education	0.024	-0.015	0.078*	0.031	0.014	0.022
Control group × Father highest education	0.081**	0.005	0.105***	0.076	0.328**	0.104**
Treatment × Father highest education	0.052	-0.001	0.197**	0.171***	0.038	0.094*
Control group × Marital status	-0.214***	0.045	-0.079	-0.015	-0.153**	-0.079
Treatment × Marital status	-0.017	0.019	0.002	0.010	-0.088	-0.021
Control group × HH asset index	-0.040**	-0.019**	-0.020	-0.056	-0.036	-0.038
Treatment × HH asset index	-0.034**	-0.002	-0.076*	-0.038	-0.021	-0.032*
ASQ communication at baseline	0.045					
ASQ gross motor at baseline		0.101**				
ASQ fine motor at baseline			0.096**			
ASQ problem solving at baseline				0.076**		
ASQ personal social at baseline					0.036	
Child development index at baseline						0.142***
t-test mother time investment (p-value)	0.560	0.621	0.584	0.953	0.814	0.998
t-test mother self-efficacy (p-value)	0.763	0.826	0.767	0.073	0.012	0.057
Observations	1174	1174	1174	1174	1174	1174
R^2	0.066	0.047	0.105	0.167	0.072	0.118

Notes: the table tests whether there is unobserved treatmen5585+§, -t effect on the observed parental inputs (mother time investment and mother self-efficacy) as explained in Appendix C.6. All estimates are OLS regressions. The table reports coefficients from a specification that regresses the child development outcomes on the interaction between the treatment status (LT=0 and FT=0 or LT and FT=1) and maternal inputs or the baseline covariates. These are child gender, child age, the total number of children in the household, the caregiver's age, the caregiver's education level, the caregiver's marital status, and the asset index. These variables are described in Appendix A.4. Mother time investment index self-reported by the mother of the child is calculated by taking the average of the three HOME-SF z-scores. Mother self-efficacy index is calculated by taking the average of the eight TOPSE z-scores. All regressions include sampling weights. The dependent variables in columns 1 to 5 include standardized z-scores of the five ASQ dimensions calculated by subtracting the control group mean and dividing by the control group standard deviation in each survey wave. The dependent variable in column 6 is the child development index calculated by taking the average of the five ASQ z-scores. A full description of the construction of the outcomes is in Section 3 and in Appendix A.3. Treatment is a binary variable equal to 1 if the respondent (the caregiver) participated in either the light treatment group or the full treatment group. Control is a binary variable equal to 1 if the respondent (the caregiver) did not participate in the treatment.
*p <10%, **p <5%, ***p <1%. Robust standard errors in parentheses are clustered at the sector level. P-values of a t-test for the null hypothesis of no difference between the interaction terms 'Control group X Mother Time Investment' and 'Treatment X Mother Time Investment', and 'Control group X Mother self-efficacy' and 'Treatment X Mother self-efficacy' are reported at the bottom of the table along with the observations and the R squared.

Source: authors' calculations based on data from the follow-up survey (2018).