



ENDLINE EVALUATION FIRST STEPS INTERA ZA MBERE PROGRAM



Save the Children®

IN PARTNERSHIP WITH

laterite
DATA | RESEARCH | ADVISORY

December 2021
Kigali, Rwanda

Authors & Contributors

Primary authors

Pietro Franchi, Laterite, pfranchi@laterite.com ; Laura Le Saux, Laterite Ltd, llesaux@laterite.com ; Magdalena Vinyard, Laterite Ltd, mvinyard@laterite.com

Cover photo: Pascal Nijibikora, Ruhango, 2021

Contributors: Eleanor Hartzell, Jean de Dieu Harerimana, Noella Kabarungi and Marie Diane Uwamahoro

Acknowledgements: This report was written by Pietro Franchi and Laura Le Saux, under the direction of Magdalena Vinyard and with editorial support from Kristin Colenbrander. Lydie Shima, Jean-Baptiste Muhoza and Linda Uwizera Mihigo led the enumerator team responsible for the endline data collection, with the support of Juliet Kalimba and Christine Nishimwe who trained and supervised the MDAT assessors. This report builds on the baseline analysis, led by Priscah Chemeli Cheruiyot, Laura Le Saux and Magdalena Vinyard with the support of Carlo Menon.

Laterite benefited tremendously from the input into the design and analysis of Professor Pasco Fearon of University College London, and from the advice on implementation and baseline analysis of the MDAT from Dr. Melissa Gladstone of the University of Liverpool, Jaya Chandna of the London School of Tropical Medicine and Hygiene, and Gareth McCray of Keele University. Laterite is grateful for the thoughtful input from Monique Abimpaye, Jean de Dieu Harerimana, Eleanor Hartzell, Noella Kabarungi and Diane Uwamahoro, from Save the Children. Laterite would also like to thank all the local leaders and survey respondents for their patience and willingness to participate in the study.

Published by

Save the Children International-Rwanda and Burundi Country Office

P.O. Box 2953, Kigali, Plot 204 | KG 9 AV, #23 Nyarutarama-Remera-Gasabo, Kigali-Rwanda

<https://rwanda.savethechildren.net/>

December 2021

Suggested citation

Franchi, P., Le Saux, L., Vinyard, M. (2021), *Final Report for the Impact Evaluation of the Save the Children First Steps program in Rwanda*. Kigali, Rwanda.

© Save the Children International 2021.

You may copy, distribute, display, download and otherwise freely deal with this work for any purpose, provided that you attribute the Save the Children International as the owner.

Disclaimer

This publication does not necessarily reflect the policy position of Save the Children International or any Save the Children Member organization. The information in this publication was based on available information at the time of preparation. No responsibility is accepted by Save the Children International or any Save the Children Member organization for any errors or omissions contained within this publication.

Contents

Acronyms.....	3
Executive Summary.....	4
Introduction & Project Background.....	7
Study Purpose & Scope.....	8
Methodology & Limitations.....	10
Study design.....	10
Data Sources.....	11
Data Analysis.....	12
Ethics & Accountability.....	12
Limitations.....	13
Findings.....	16
Demographic data & respondent characteristics.....	16
Effects of the First Steps intervention on caregiver practices.....	17
Effects of the First Steps intervention on developmental outcomes.....	24
Conclusions.....	27
Recommendations for management action.....	28
References.....	29
Appendix I: Indicator Tables.....	30
Appendix II: Sample Size Calculations.....	34
Appendix III: Quality of the MDAT administration.....	35
Appendix IV: Robustness Checks.....	38
Analysis of the Average Treatment effect on the Treated (ATT).....	38
Analysis of the Local Average Treatment Effect (LATE).....	38
Heterogeneous Treatment Effects.....	39
Appendix V: Terms of Reference.....	40
Appendix VI: Data Collection Instruments.....	40

Acronyms

ANCOVA	Analysis of covariance regression model
ASQ	Ages & Stages Questionnaires
ATT	Average Treatment effect on the Treated
ECD	Early Childhood Development
ITT	Intention to Treat
IZU	Inshuti Z'Umuryango
KAP	Knowledge, Attitudes, and Practices
LATE	Local Average Treatment Effect
MDAT	Malawi Developmental Assessment Tool
RCT	Randomized Controlled Trial

Executive Summary

Project Background

Save the Children's First Steps "Intera za Mbere" parenting program aims to improve parenting practices, child development outcomes, and promotion of emergent literacy in the home. The program targets caregivers of children under three years of age. The program is delivered through community-based parenting education sessions, home activities, and radio programs in Rwanda focusing on four areas: responsive caring, playful learning, language & literacy, and healthy beginnings. The program was implemented in partnership with a local organization, Umuhuza, in two cohorts in Gasabo, Kirehe and Ruhango districts. The first cohort, which is the focus of this evaluation, received the intervention between November 2019 and March 2020. The second cohort received the intervention between November 2020 and March 2021 with an adaptation of the First Steps package due to COVID-19 pandemic.

First Steps "Intera za Mbere" is delivered in two different versions: full and light-touch. Both versions include 18 parenting education sessions delivered weekly in the community by *Friends of Families* - Inshuti z'umuryango (IZU) using a radio program and providing community-based activities such as talking, singing, reading, playing, bonding, providing advice to pregnant women, and guidance to parents on nutrition and other health topics for the development of children. Families receiving the full version also received at least one home visit of their IZU during the intervention; vulnerable families received additional home visits from IZU as needed. In addition, families in the full version received materials and access to a children's book bank.

Study Purpose and Key Questions

This evaluation is a cluster Randomized Controlled Trial (RCT) conducted in the districts of Kirehe, Ruhango and Gasabo. It is designed to evaluate the two versions, full and light-touch, of the First Steps "Intera za Mbere" intervention against a control group to determine causal effects.

This evaluation explores changes in caregiver practices and developmental outcomes that resulted from the intervention. This study aims to evaluate the comparative impact of the two versions of the intervention on (i) caregiver knowledge, attitudes and practices related to early childhood care and development and (ii) children's developmental outcomes. In this report, we only discuss statistically significant findings.

Conclusions

Both full and light-touch versions of the First Steps "Intera za Mbere" parenting program improved self-reported responsive parenting practices. The availability of books in the households was the only aspect more strongly impacted by the full version than the light-touch version.

- Households that received either version of the intervention increased the variety of interactions caregivers engage in with their children compared to the control households.
 - The intervention increased the share of caregivers who engage in interactions with their child like reading, telling stories, singing songs, and naming, counting, and drawing things with the child. In addition, the light-touch intervention increased the number of caregivers who play and go outside with the child.
 - The intervention increased the diversity of interactions caregivers engage in with their child. Caregivers who received either the full or light-touch intervention reported that their child was engaged in an average of four different types of interactions in the last three days, compared to three for the control group.
 - The intervention increased the involvement of fathers in parenting practices, with a larger increase in the share of fathers engaging in at least one type of interaction with their child in the last three days in the intervention groups (+8 percentage points (pp) for the full treatment and +13 pp for the light-touch version).
- The intervention increased book availability in the household, with a stronger effect for households who were exposed to the full version. Both the share of children who have access to books and the average number of

books in the household increased as a result of the intervention. Children from the full treatment group have two books on average, compared to one in the light-touch group.

- **The share of caregivers leaving their child alone for more than an hour increased less in treatment groups than in the control group.** The effect was larger for households assigned to the full intervention (+1 pp compared to +2 pp for light-touch households and +7pp for control households). However, the intervention did not impact the proportion of children left in the care of another child for more than one hour.

Both the full and light-touch intervention arms improved positive discipline practices.

- **The intervention reduced the use of physical punishment.** Prevalence of these violent methods decreased by 15 and 16pp respectively in the light-touch and full treatment groups, compared to 7pp in the control group.
- **The intervention increased the use of non-violent discipline methods.** Prevalence of these methods increased by 35 and 36pp respectively in the full and light-touch groups, compared to 19pp in the control group.
- **However, the intervention did not affect attitudes towards physical punishment, and did not reduce the use of psychological aggressions.**

The effects of the intervention on health and nutrition practices are more limited:

- Caregivers in the light-touch group are significantly more likely to meet the minimum meal frequency requirements when feeding their child than control caregivers.
- The intervention did not reduce the incidence of diarrhoea, fever, cough or difficulty breathing.

Some responsive parenting practices are associated with positive change in developmental outcomes and suggest that the improvement in caregiver practices could eventually lead to improvement in developmental outcomes. The diversity of interactions caregivers engage in and the use of non-violent discipline methods (as opposed to physical and psychological violence) are associated with improved developmental trajectories. These key practices were also improved by the intervention, indicating a mechanism through which the First Steps “Intera za Mbere” program likely impacts the trajectory of developmental outcomes.

Neither version of the intervention had a significant impact on children’s developmental trajectory as measured by the Malawi Developmental Assessment Tool (MDAT). However, this does not necessarily mean that the intervention did not or will not impact developmental outcomes: these effects may take more time to materialize, or the tool may fail to capture them.

Three strong drivers of the trajectory of development of children are the caregiver’s educational attainment, the child’s gender and whether the child has a disability. At endline, like at baseline, girls scored higher than boys, and children with disabilities scored lower than children who did not have disabilities. Those characteristics also predict differentiated developmental trajectories between baseline and endline: girls developed skills faster than boys, while children with a disability developed measured skills more slowly than children who did not have a disability. Similarly, children whose primary caregiver had completed secondary education developed cognitive skills faster than other children.

Recommendations for data use

The evaluation confirms that the First Steps “Intera za Mbere” program positively impacts a range of responsive parenting, early learning, positive discipline and health and nutrition practices and, through them, likely contributes to positive developmental changes. This evaluation confirms the findings of the pilot study in demonstrating the impact of the intervention on a variety of caregiver practices. In particular, the intervention led to improvements in the diversity of interactions caregivers engaged in with their children and in the use of positive discipline methods, which in turn are positively associated with children’s developmental trajectories. This signals a mechanism through which the First Steps “Intera za Mbere” program likely contributes to positive change in children’s developmental trajectories.

Overall, the full version of the intervention does not lead to larger improvements than the light-touch version. The evaluation did not find a systematically stronger impact of the full intervention on parenting practices, which suggests that

the additional visits of the IZUs to the households did not lead to substantial additional changes in behavior. However, the full version of the intervention did increase the availability of books to a larger extent than the light-touch version, as expected since the latter did not include any provision of material promoting literacy. While in some cases the impact of the light-touch intervention seems higher than for the full version, these differences are not significant.

The evaluation highlights opportunities to increase emphasis on several topics within the program. In particular, while the intervention led to increased use of non-violent discipline methods and decreased use of physical violence, it did not reduce the use of psychological aggressions. We also found limited effects on prevalence of disease and related care-seeking behaviors, suggesting that best practices in this matter could require emphasis within sensitization sessions.

Methodology and Limitations

This evaluation uses quantitative data collected in September 2019 (baseline) and in September 2021 (endline). The endline data collection, initially planned in May 2020, was shifted to September-October 2021 as a consequence of the COVID-19 pandemic. Of the 1,184 households included in the study at baseline, 1,107 were interviewed at endline.

The conclusions of this evaluation draw from an Intention to Treat analysis. The analysis is based on the estimation of an ANCOVA model, which focuses on explaining the change in the outcome value (developmental score or caregiver practice) by including the initial (baseline) value of this outcome together with the treatment status. Our analysis focuses on the impact of the intervention on caregivers and children that were meant to receive it, whatever their actual behavior and treatment status.

We do not know the extent to which treatment households were actually exposed to the version of the intervention they were assigned to. Our assessment of households' compliance to the treatment group they were assigned to is based on self-report 18 months after the end of the intervention and is thus imperfect. A limited number of control group caregivers reported listening to radio programs, attending parenting sessions or receiving visits from an IZU, suggesting low contamination of this counterfactual group. Reported attendance of parenting sessions was high for both light-touch and full treatment groups. However, less than 50% of full treatment caregivers recalled a visit from an IZU in relation with the program and 20% of light-touch caregivers reported at least one. This might contribute to explain the limited differences we found across the two versions of the intervention.

Introduction & Project Background

Save the Children's First Steps "Intera za Mbere" parenting program aims to improve parenting practices, child development outcomes, and promotion of emergent literacy in the home. The program targets caregivers of children under three years of age and is implemented in partnership with a local organization, Umuhuza.

The First Steps "Intera za Mbere" program was implemented in Gasabo, Kirehe and Ruhango in two cohorts. A first cohort of 2,421 families received the intervention between November 2019 and March 2020. Between November 2020 and March 2021, a second cohort of 2,257 families received a version of the intervention that was adapted to the COVID-19 pandemic. This evaluation was conducted only with households in the first cohort.

First Steps "Intera za Mbere" is delivered through parenting education group sessions using blended local radio programming and community-based activities. The blended approach allows the program to reach more families at lower cost without compromising quality. Parenting sessions are delivered weekly in the community by trained Inshuti z'umuryango (IZU)¹ across four focus areas: responsive caring, playful learning, language & literacy, and healthy beginnings. Caregivers listened to the radio program during parenting sessions.² In addition, all trained IZUs receive a facilitator's guide, illustrated posters and booklets to support the sessions of this first cohort.³ The parenting education curriculum includes 18 in-person sessions of 75 to 90 minutes:

- | | |
|--------------------------------|--|
| 1. Responding and bonding | 10. Developing through play |
| 2. Playing with a child | 11. Early language and communication |
| 3. Talk, read and talk | 12. How you promote literacy |
| 4. Feeding with care | 13. Book making |
| 5. Calming and soothing | 14. Literacy alive in home and community |
| 6. Routines | 15. Partners for a healthy pregnancy |
| 7. Positive discipline | 16. Healthy baby: pound of prevention |
| 8. Materials at home for play | 17. Health clinic now |
| 9. Creative play – make a book | 18. Cooking demonstration |

The program was implemented in two different versions: full and light-touch. Both versions received the parenting education sessions as described above. Families receiving the full version also received at least one home visit during the intervention; vulnerable families received additional home visits from IZUs as needed. In addition, families in the full version received materials and access to a children's book bank. The theory of change of the intervention, shown in Figure 1, illustrates the sequence of inputs, outputs, and outcomes of the intervention.

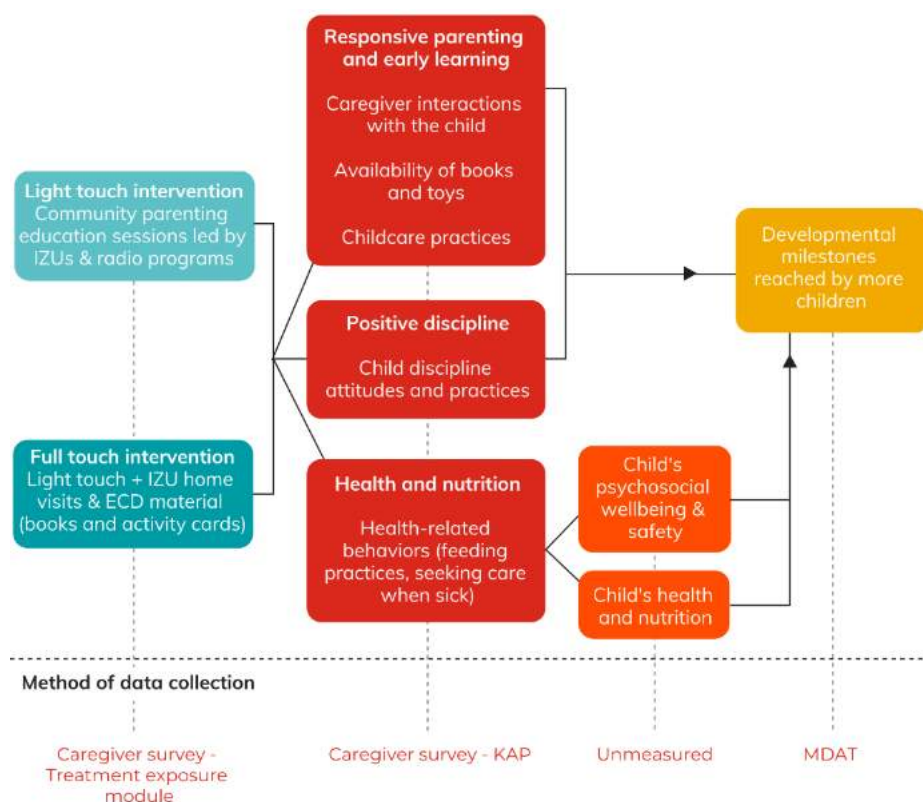
In response to the COVID-19 pandemic, Save the Children adapted the First Steps "Intera za Mbere" program implemented with the second cohort and broadcasted radio programs nationally. Due to the COVID-19 pandemic, the radio program was adapted by including messages on COVID, caregiver well-being, caregiver-child interactions, and maintaining, promoting nurturing childcare during COVID-19, positive discipline, and male engagement during and beyond the lockdown period. The radio program was complemented by phone calls from community volunteers to remind caregivers to listen to the radio show, and home visits conducted by the community volunteers to emphasize the messages that they have listened to the radio program.

¹ Since 2015, one man and one woman are elected in each village to serve as community-based child and family protection volunteers. Their main responsibilities are to promote child rights, protect children from violence, abuse, and exploitation, mobilize against early pregnancies and school dropouts, and promote equal rights of children with disabilities. They report to their respective village chief/leader and to their representative in the cell and sector, up to the district, which follow up on all reports for an inclusive child protection system.

² Radio programs were broadcasted locally and were not advertised. Treatment groups caregivers were directly exposed to them during group sessions. This is different from the potential more passive exposure of control group caregivers who may have heard these programs by chance.

³ Following the COVID-19 pandemic, the program was adapted in several ways. In particular, volunteers do no longer receive and use posters.

Figure 1: Theory of change of the First Steps “Intera za Mbere” intervention



Study Purpose & Scope

Context & Study Purpose

This intervention draws on the nurturing care framework for early childhood development which includes good health, adequate nutrition, responsive caregiving, security and safety, and opportunities for early learning.⁴ The nurturing care framework takes a holistic view of the influences on child development including the role of policies, programs, systems, communities, and individuals. The First Steps “Intera za Mbere” intervention, however, focuses primarily on the role of caregivers and nurturing caregiving in promoting child development. Nurturing caregiving includes knowledge, attitudes, and behaviors around caregiving activities such as maintaining the health and hygiene of children and feeding practices; stimulation such as talking and playing with children; responsiveness including responding to children’s interests and needs, creating trust and attachment, and sensitive communication; and providing children with safety and protection from harm.⁵ These core components are therefore reflected in the focus of this evaluation.

This evaluation aims to build on the results of a 2015-2016 evaluation of the pilot phase of the First Steps “Intera za Mbere” intervention. The pilot evaluation in Ngororero district found that both the full and light-touch versions of the program improved child health and nutrition practices and parent-child engagement, which contributed to children reaching appropriate developmental milestones. However, during the implementation of the pilot program, the full intervention was

⁴ Reference: World Health Organization, United Nations Children’s Fund, World Bank Group. Nurturing care for early childhood development: a framework for helping children survive and thrive to transform health and human potential. Geneva: World Health Organization; 2018. Licence: CC BY-NC-SA 3.0 IGO.

⁵ Reference: Britto PR, Lye SJ, Proulx K, Yousafzai AK, Matthews SG, Vaivada T, Perez-Escamilla R, Rao N, Ip P, Fernald LCH, MacMillan H, Hanson M, Wachs TD, Yao H, Yoshikawa H, Cerezo A, Leckman JF, Bhutta ZA; Early Childhood Development Interventions Review Group, for the Lancet Early Childhood Development Series Steering Committee. Nurturing care: promoting early childhood development. Lancet. 2017 Jan 7;389(10064):91-102. doi: 10.1016/S0140-6736(16)31390-3. Epub 2016 Oct 4. PMID: 27717615.

closer to the light-touch than intended as there were delays in implementing some components of the full intervention for the first cohort. Following this pilot, Save the Children scaled up the project to three districts with a specific focus on early childhood development components of the intervention.

The program's learning objectives are to explore the comparative impact and scalability of the two versions of the intervention, full versus light-touch. Scaling up this intervention was an opportunity to extend the benefits of First Steps to more parents and children in Rwanda while testing the relative benefits of each of the two treatment groups with a larger sample of parents. Randomization of areas to treatment and control is done at the cell rather than sector level, which means that this evaluation has more statistical power to detect an effect between treatment groups than the pilot evaluation. These results will inform the scaling of the program as well as Save the Children's advocacy efforts at the national level.

Baseline findings

The baseline report for this study⁶ indicates that some of the positive parenting practices promoted by the First Steps “Intera za Mbere” intervention are already common. For example, for most households, breastfeeding practices are in line with UNICEF and World Health Organization recommendations; most children play with some kind of toy; and, some types of caregiver interactions with the child such as singing, playing, and going outside are widespread.

However, the baseline results also show an opportunity for improvement in many other parenting practices First Steps “Intera za Mbere” targets. These include promoting a variety of types of caregiver-child interactions; emphasizing the need for stimulation with children of all ages including very young; increasing the variety in interactions between fathers and their children; increasing the availability and use of books for young children and toys in the home; increasing health care-seeking behavior within the formal (public, private, or community-based) health systems; increasing awareness and use of positive discipline; and reduction of physical violence as a form of child discipline.

The results from the baseline also show associations between key links in the theory of change, providing some support for the proposed mechanism of the intervention. Two key intermediate outcomes – increased caregiver interactions of all types and availability of books and toys - are strongly associated with higher child development scores at baseline. The baseline results also show that higher exposure to parenting messages (prior to the intervention) is associated with a higher prevalence of related caregiving practices. This supports the theory of change that a radio blended model may influence behavior change in parenting practices.

Study Questions

This evaluation aims to assess the impact of the full and light-touch interventions on developmental outcomes, our main study question. According to the theory of change, these hypotheses are expected to hold provided that the full and light-touch interventions affected caregiver knowledge, attitudes and practices related to early childhood care and development, our secondary study question. For both study questions, we formulate three research questions of interest:

1. Does the full intervention lead to significant improvements compared to the control group?
2. Does the light-touch intervention lead to significant improvements compared to the control group?
3. Does the full intervention lead to significantly larger improvements than the light-touch intervention?

⁶ Laterite, in collaboration with Save the Children, published a baseline report that illustrates the preliminary findings from baseline data collection (Laterite, 2020).

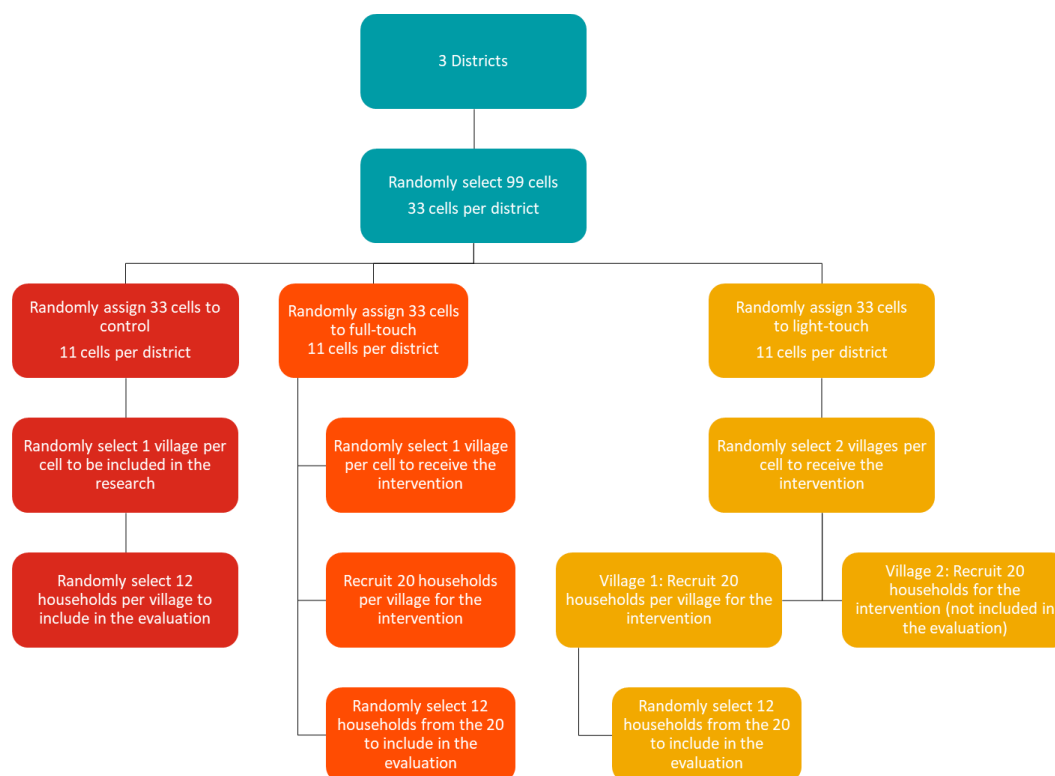
Methodology & Limitations

Study design

This evaluation is a cluster Randomized Controlled Trial (RCT), designed to evaluate two forms of the First Steps “Intera za Mbere” intervention, full and light-touch, against a control group. The baseline survey was conducted in September 2019, before the start of the intervention, and the endline, initially planned in May 2020, was shifted to September-October 2021 as a consequence of the COVID-19 pandemic.

This evaluation took place in three districts and is clustered at the cell level. This evaluation took place in three districts: Gasabo, Ruhango, and Kirehe.⁷ The intervention was randomly assigned at cell level, with full intervention, light-touch, and control cells in all districts and sectors. In each cell, one village or a pair of villages⁸ was randomly selected for inclusion in the evaluation. In each village, 20 households with a child between the ages of 6 and 24 months were randomly selected for inclusion in the intervention. Of these, 12 were randomly selected for inclusion in the evaluation. The design is summarized in Figure 2.

Figure 2 : Intervention and research design of First Steps “Intera za Mbere”



Sampling Methods & Sample Size

Randomization and stratification at the cell level ensured satisfactory statistical power while limiting the risk of spillover. The selection of cells into the trial was stratified at the district level: 33 cells (clusters in the RCT) were randomly selected for inclusion in the evaluation in each of the three districts. By selecting only one (or in select cases two) village(s)

⁷ Within these districts some areas are excluded. Three sectors in Gasabo, which have already received the intervention, one sector in Kirehe, which does not have access to radio, three sectors and four cells in Ruhango, which already benefit from other early childhood development interventions.

⁸ There are between three and 25 villages per cell. Within each cell, one village was randomly selected to participate in the trial. In 25% of cells, there were not enough eligible households in the first village selected, and a second adjacent village was included in the evaluation as well.

per cell, we aimed to limit the likelihood of communication between households enrolled in different treatment arms of the trial (spillover effects). Such spillover effects between clusters could impact the integrity of the evaluation.⁹

A household was eligible to participate in the trial at baseline if they had a child aged 6 to 24 months at the time of household listing. Within each village selected for the trial, Laterite listed every household with children under 3, and checked for their eligibility. Laterite randomly assigned every eligible household in the selected village(s) a rank and selected the first 12 as main households with the remaining being replacements. However, all caregivers and child pairs in a selected household were invited to participate in the intervention. A primary caregiver was defined as whoever the family members self-report as the person with primary responsibility to attend to the bodily, health, nutrition, emotional, social, language and intellectual development needs of a child, including a parent, family member, or other person accorded with such duties.

At endline, we aimed to enroll all baseline households that could be located and were still eligible for the study. Households were not included in the endline survey if they had moved out of the study districts of Kirehe and Ruhango or out of the Kigali province (which contains Gasabo district). Households were also excluded if the eligible child in the household had passed away or if the caregiver and child did not consent or were not available for the whole data collection period. Out of the baseline sample of 1,184 households, 77 were dropped at endline (12 could not be located and 65 were not eligible).¹⁰

Data Sources

This evaluation uses quantitative data collected in September 2019 (baseline) and September 2021 (endline) in Kirehe, Ruhango and Kigali. The enumerator team was responsible for administering a 30-minute caregiver survey and a 30-60-minute child development assessment in Kinyarwanda through SurveyCTO. Because of the complexity of administering the child assessment, which requires a semi-private space and considerable equipment, data collection occurred at one central location per village (or pair of villages) rather than at each household. This was not only a more cost-effective and logistically feasible way to administer the tool, but also allowed for more supervision of the enumerator team, increasing reliability and consistency among assessments.

Caregiver knowledge, attitudes and practices were assessed through a caregiver survey. At endline, the caregiver interview included modules related to (i) caregiver & household characteristics, (ii) the household's experience during the COVID-19 pandemic, (iii) Knowledge, Attitudes, and Practices (KAP) towards early childhood development, especially regarding health and nutrition, responsive parenting, early learning and positive discipline, and (iv) exposure to the Intera za Mbere program.

Child development was assessed using the Malawi Developmental Assessment Tool (MDAT). The MDAT relies on observations of a child as they complete a variety of tasks, supplemented by caregiver-report, to measure development across four domains: gross motor skills, fine motor skills, language skills, and social skills. The tool is appropriate for use with children 0-5 years of age and is appropriate for use in low-resource contexts (see Boggs et al. 2019). Given the complexity of the MDAT, Laterite assessed reliability of administration during a small pilot prior to data collection by having two enumerators simultaneously observe and score child assessments. Details of the pilot protocols and results of the inter-rater reliability calculation at endline can be found in Appendix III.

The urbanization status of each study sector was computed using the fourth Population and Housing Census data (2012).

⁹ We could have further reduced the possibility of spillover by randomizing the trial at a higher administrative unit, such as sector. However, there are only a small number of sectors in each district compared to cells (27 sectors across three districts, compared to 99 cells). This approach would have decreased the number of clusters, and therefore the statistical power of the trial. In the spillover/power trade-off, we favored statistical power over marginal reductions in spillover.

¹⁰ 6 children passed away, 41 relocated outside of the study areas, 2 did not consent and 15 were unavailable for the data collection period. Additionally, among the 1,108 households that we could reach, one child was sick and could not be assessed.

Data Analysis

Our analysis builds on the study design to ensure generalizability of the results. The statistical analysis was conducted through Stata version 15 using weights to consider the two-stage sampling and clustered design of the study, as well as attrition (see more details in the Limitations section). By weighting the survey data, we can generalize the results of the study to the population from which it was sampled - that is, the population of the three districts not including areas excluded from the sample.

In this endline report we estimate causal impact from the intervention using the Intention to Treat (ITT) estimator. The ITT analysis focuses on the impact of the intervention on caregivers and children that were meant to receive it, whatever their actual behavior and treatment status. For various reasons, caregivers might not have complied with their original treatment assignment (see more details on compliance in the Limitations section). Other estimators (e.g., ATT, LATE) allow one to account for this by estimating the impact of the intervention on caregivers and children who did comply with their original treatment status. In this study, exposure to the intervention was captured using caregiver self-report, which is an imperfect measure of compliance, especially more than 18 months after the end of the intervention. This explains why our main estimator is the ITT, which is furthermore better suited to inform a possible scaling-up of the intervention.

The analysis is based on the estimation of an ANCOVA model. This model focuses on explaining the change in the outcome value (MDAT score or caregiver practice) by including the baseline value of this outcome, together with the treatment status (full or light-touch) and control variables observed at baseline (here, child gender, age¹¹ and disability status,¹² caregiver educational attainment and marital status, Ubudehe category, sector urbanization status).

Descriptive statistics about the population surveyed, and associations between key variables in the survey were also explored. We tested if differences in key outcome variables between groups with different characteristics were statistically significant. Whenever the key outcome was a categorical variable (e.g., the availability of mosquito nets in the household, whether the caregiver played with the child within the three days preceding the survey), we used a Pearson Chi-squared test to detect statistically significant associations. Whenever the key outcome was a continuous variable (e.g., MDAT scores), we ran univariate linear Ordinary Least Squares (OLS) regressions using either a household/caregiver/child characteristic or a caregiver practice as the explanatory variable to detect statistically significant associations. Only statistically significant associations are discussed throughout the report.

Ethics & Accountability

Research Ethics

Consent was collected from adult participants for both the caregiver and the child participating in the study. When recontacting households, Laterite verbally requested consent from respondents to participate in the study. On the day of the survey, we requested written informed consent from all caregivers who we interviewed for their participation and that of their child. The enumerator read aloud the consent script in Kinyarwanda to the respondent. Participants were given a chance to ask questions and then accept or decline to participate; however, no one declined. To give their written consent, participants signed the consent form electronically on a tablet. Participants were given a paper copy of the consent form for their records.

This research received approvals from relevant Rwandan Institution Review Boards. This research received approvals from the Rwanda National Ethics Committee (N°621/2019, N°922/2020, N°167/2021 and N°825/2021), National Institute of Statistics Rwanda (N°348/2019, N°254/2021, N°287/2021) and National Council for Science and Technology (482/130/2019, 482/236/2021).

¹¹ The child age is only included when the outcome is a caregiver practice, while MDAT scores are age-standardized.

¹² Note that we included the reported disability status at endline rather than baseline.

Limitations

The main limitations of this study relate to internal validity and include attrition, contamination and non-compliance risks.

Attrition is limited and was accounted for in the analysis. We managed to find 94% of households at endline. The remaining 6% were dropped either because they could not be reached or because they were not eligible for the study anymore.¹³ This is a relatively small attrition rate which falls within the sampling and attrition assumptions made at baseline, so that it does not affect our overall ability to detect impact in this study (limited loss in statistical power).¹⁴ Attrition is associated with the treatment status (full treatment households are more likely to have dropped from our sample), residence district, and unemployment status and educational attainment of the caregiver (see Table 1). Weights were adjusted to account for the change in representativeness of the sample compared to baseline.¹⁵

Table 1: Characteristics across households that left the sample and households that were present at endline

	Dropped households	Endline Households	Difference
District			
Gasabo	57%	32%	25pp***
Kirehe	30%	34%	-3pp
Ruhango	43%	68%	ref
Caregiver's characteristics			
Partnered	79%	81%	-2pp
Non-partnered	21%	19%	ref
Primary educ.	51%	50%	2pp
Secondary educ. and more	36%	19%	16pp**
No school	13%	31%	ref
Self-employed	12%	5%	7pp
Farmer	23%	55%	-33pp***
Daily wages worker (agric. sector)	29%	21%	8pp
Unemployed	12%	7%	6pp
Other employment	24%	12%	ref
Socio-economic status			
Ubudehe 1	11%	16%	-5pp
Ubudehe 2	55%	50%	5pp
Ubudehe 3	34%	33%	ref
Child characteristics			
Child has at least one type of disability	16%	12%	4pp
BL Child age in months	16.07	15.13	0.94
Male child	58%	54%	4pp
Child is covered by health insurance	66%	75%	-9pp
Urban/rural status			
Rural	50%	43%	7pp
Urban	38%	52%	-13pp
Mixed	12%	5%	ref
Treatment status			
Full treatment	47%	32%	15pp*
Light touch	26%	34%	-8pp
Control	26%	34%	ref

Note: ***, **, and * indicate significance at the 0.1, 1, and 5 percent critical level. Proportions are from the sample before any sampling weights were applied, therefore endline characteristics used in the rest of the report are adjusted to be representative of the population

¹³ Either because they relocated outside of the study areas, they were unavailable for the whole data collection period or the child passed away.

¹⁴ Study design included a 10% attrition assumption.

¹⁵ We applied propensity score weights equal to the inverse of the probability for a household to be interviewed at endline. This probability is predicted by a logistic regression using as independent variables the baseline values of the household district, their urban/rural status, the level of education of the caregiver and the caregiver occupation. These weights were then combined with baseline survey weights, so that the final sample is both corrected for non-random attrition, and representative of the population of interest. When building propensity score weights, we used multiple imputation through multinomial logistic regression to predict the two baseline caregiver educational attainment values that were missing.

of interest. For each category, “ref” indicates the reference category for which no difference test was conducted. If each variable has n categories, we only need to test the difference of $(n-1)$ of them, since the n th category is a linear combination of the others.

Contamination through recent radio program broadcasting is limited. Contamination refers to the transmission of the effects of the intervention from the treatment group to the control group. Due to the national broadcasting of the Intera za Mbere radio program to respond to the COVID-19 pandemic, the study control group was potentially exposed to radio sessions between November 2020 and March 2021, which meant that control group households may have benefited from the intervention to a larger extent than initially intended.¹⁶ To assess spillovers in the study, questions were added to the treatment exposure module of the caregiver survey to ask about exposure to the radio program. This approach remains imperfect since it relies on respondents recalling specific radio programs heard since November 2020. A minority of the control group caregivers (7%) reported that they listened to those programs within the past year, compared to respectively 37% and 35% of the full and light-touch group caregivers (see Table 2). This means that this channel of contamination was limited.

The control group reported limited intervention take-up (see Table 2). Among caregivers we re-interviewed¹⁷, a minority of caregivers from the control group recalled participating in Intera za Mbere parenting education sessions (8%) and receiving a visit from an IZU in relation to the Intera za Mbere program (4%). Overall, 9% of the control group caregivers reported being exposed to the intervention to some extent (indicating that compliance with their assignment to control group was high). Contamination seems all the more limited given that control group caregivers reported attending three parenting sessions (out of 18) on average, compared to 12 for both treatment group caregivers.

Table 2: Self-reported exposure to the First Steps “Intera za Mbere” program

	Control	Light	Full	Total	Difference
Exposed to cohort 2 radio programs	7%	35%	37%	26%	***
Participated in at least on parenting session	8%	83%	88%	60%	***
Received an IZU visit in relation with the intervention	4%	20%	47%	23%	***

Note: ***, **, and * indicate significance at the 0.1, 1 and 5 percent critical level.

Non-compliance seems more important across treatment groups. Non-compliance refers to the individuals not complying with their treatment assignment. A majority of the full and light-touch groups (88% and 83% respectively) attended the sessions, which shows high take-up. However, only 47% of caregivers from the full treatment group recalled receiving a visit from the IZU, compared to 20% of the light-touch group, which shows both low take-up and substantial contamination. There is no difference in the average number of reported IZU visits across those two groups (on average, 1.8 visits in the light-touch group and 1.7 visits in the full treatment group). However, among caregivers who reported receiving an IZU visit in relation with the program, full treatment caregivers were more likely to recall receiving printed materials (books, home cards) from the program: 96% did, compared to 37% of light-touch caregivers who reported an IZU visit in relation with the program.

Our analysis shows that non-compliance is substantial in this study, although it may be affected by memory biases. More investigation is needed to determine whether caregivers misreported their exposure to the intervention or if they were potentially (un-)reached by IZUs by mistake. In the absence of monitoring data, our Average Treatment effect on the Treated (ATT) analysis, which serves as a robustness check (see Appendix IV), should be taken with caution as self-reported exposure to the intervention is likely to be affected by various memory biases.

Our analysis of caregiver behavioural changes relies on self-reported practices. The parenting practices that we use as outcomes of our analysis are based on caregiver self-report, which might suffer social desirability bias, which describes

¹⁶ It should be noted that the national broadcasting of the radio program took place after the first cohort completed the intervention as originally planned and that exposure to the radio program broadcast nationally cannot be considered comparable to exposure as part of the intervention. In both the full intervention and in the light touch version, caregivers listened to the radio program during parenting sessions and were therefore directly exposed to it. This is different from the more passive exposure of the intervention broadcast nationally – neither treatment nor control households were notified about the national broadcast of the intervention.

¹⁷ In 6% of households, the primary caregiver of the child had changed since baseline and we interviewed the new caregiver.



the tendency of survey respondents to answer questions in a manner that will be viewed favourably by others. While measured parenting practices were similar across groups at baseline, this bias might have increased for caregivers who were exposed to the intervention. Similarly, the child disability status relies on their caregiver's judgement.

For this study, the main risk to external validity relates to the geographical scope of the study. Extrapolating the results to other districts – especially those that are not comparable in terms of households' socio-economic characteristics and caregivers' attitudes and practices – requires some caution. However, it is worth mentioning that the areas included in the study were chosen because they had no other interventions at the time of design, and this was confirmed at endline.

Findings

Demographic data & respondent characteristics

The few changes in household demographics from baseline to endline are likely to stem from the COVID-19 pandemic. We investigated the changes in households' characteristics between baseline and endline, noting that overall, demographics remain the same (see Table 3). About 48% of caregivers have completed primary education and 22% have completed secondary education or more. The average household size was around five members, and 80% of caregivers reported having a partner. Almost half (48%) of the sample was from rural areas, while 46% was from mixed areas and the remaining 6% from urban areas. 55% of the children are male, which differs slightly from the national distribution (where 50% of children are male). The distribution of caregiver occupation significantly changed from baseline to endline, with lower shares of farmers and unemployment and higher shares of daily wages employment (in both agricultural and non-agricultural sectors) at endline. In addition, the share of children covered by health insurance rose sharply, from 75% at baseline to 91% at endline. This share is comparable across treatment groups.

Table 3: Household characteristics at baseline and endline

	Baseline	Endline	Difference
District			
Gasabo	37%	38%	*
Kirehe	40%	39%	
Ruhango	23%	23%	ref
Caregiver's characteristics			
Partnered	81%	80%	
Non-partnered	19%	20%	ref
Primary	48%	48%	
Secondary and more	23%	22%	
No school	29%	30%	ref
Self-employed	7%	8%	
Farmer	48%	35%	***
Daily wages worker (agric. sector)	21%	37%	***
Unemployed	16%	8%	**
Other employment	8%	12%	ref
Ubudehe			
1	14%	13%	
2	53%	53%	
3	34%	33%	ref
Child characteristics			
Child has at least one type of disability	12%	7%	***
Male child	55%	55%	
Child is covered by health insurance	75%	91%	***
Urban/rural status			
Urban	6%	6%	
Rural	48%	48%	
Mixed	46%	47%	ref

Note: ***, **, and * indicate significance at the 0.1, 1 and 5 percent critical level.

The primary caregiver changed in a minority of households, with some differences in household characteristics. The primary caregiver changed in 7% of households between baseline and endline. A majority of these new caregivers were grandparents, and they were less likely to be partnered than the remaining primary caregivers. The distribution of educational attainment, marital status and urbanization status remained broadly unchanged.

Prevalence of child disability evolved between baseline and endline. Our measure of child disability relies on a set of screening questions included in the MDAT assessment¹⁸ that differs from the Washington group questions. The share of caregivers reporting at least one disability for their child dropped from 12% to 7% of children. Because caregivers tend to more reliably detect child disability as the child grows older, we used the endline reported disability as a control variable in the remainder of the analysis.

Box 1: General effect of COVID-19 on the household

The baseline survey was conducted in September 2019, and the intervention took place from November 2019 to March 2020, right before the COVID-19 pandemic started. This brought changes to the evaluation, with the endline survey being postponed from May 2020 to September 2021. This and following boxes sum up findings derived from the COVID-19 module of the endline survey, which included questions about the household's experience during the pandemic and any self-reported changes in general activities, income and home life.

The COVID-19 pandemic brought important economic difficulties for all groups. Respondents reported that the areas made most difficult by COVID-19 were getting food (88%), caring for vulnerable people (82%), and caring for children (70%). The worst economic effects of the pandemic were on household income (losses were reported by 87% of the sample) and on job opportunities (86% reported working less, while 85% reported finding fewer job opportunities).

Control group caregivers were more likely to report severe stress. When reporting the stress caused by COVID-19 in different areas, the control group consistently reported more extreme stress due to the pandemic, as opposed to the full intervention arm and especially the light-touch group.

Effects of the First Steps intervention on caregiver practices

1. Responsive parenting practices

The full and light-touch versions of the intervention increased caregivers' engagement with their child across almost all types of interactions. Both the full and light-touch interventions had a positive and significant impact on caregivers reading books, telling stories, singing songs or naming/counting/drawing things with their child (see Figure 3 and Figure 4). There was a positive significant increase in caregivers playing with their child and taking their child outside in the light-touch group compared to the control, but not in the full treatment group. However, the overall increase in those practices was smaller than for others, as they were already commonly reported at baseline.

Caregivers with educational attainment and from rural households were more likely to increase interactions with their child. Caregivers with secondary education were more likely to increase their engagement in some interactions between baseline and endline, i.e., reading books, naming/counting/drawing things, and playing with the child. The same is true for caregivers from rural households, who, in addition, were more likely to increase their engagement in singing songs and taking their child outside.

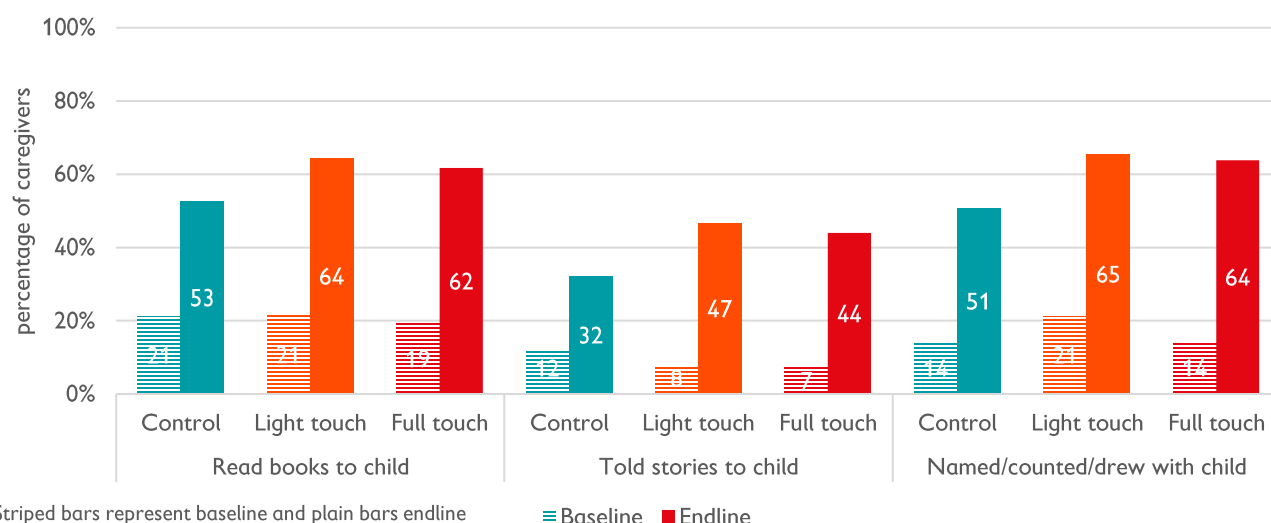
¹⁸ Compared with other children, does your child (1) have any serious delay in sitting, standing, or walking; (2) have difficulty seeing, either in the daytime or at night; (3) appear to have difficulty hearing? (uses hearing aid, hears with difficulty, completely deaf?); (4) have difficulty in walking or moving his/her arms or does he/she have weakness and/or stiffness in the arms or legs; (5) sometimes have fits, become rigid, or lose consciousness? (6) learn to do things like other children his/her age; (7) appear in any way mentally backward, dull or slow?

First Steps program evaluation

December 2021 - Rwanda

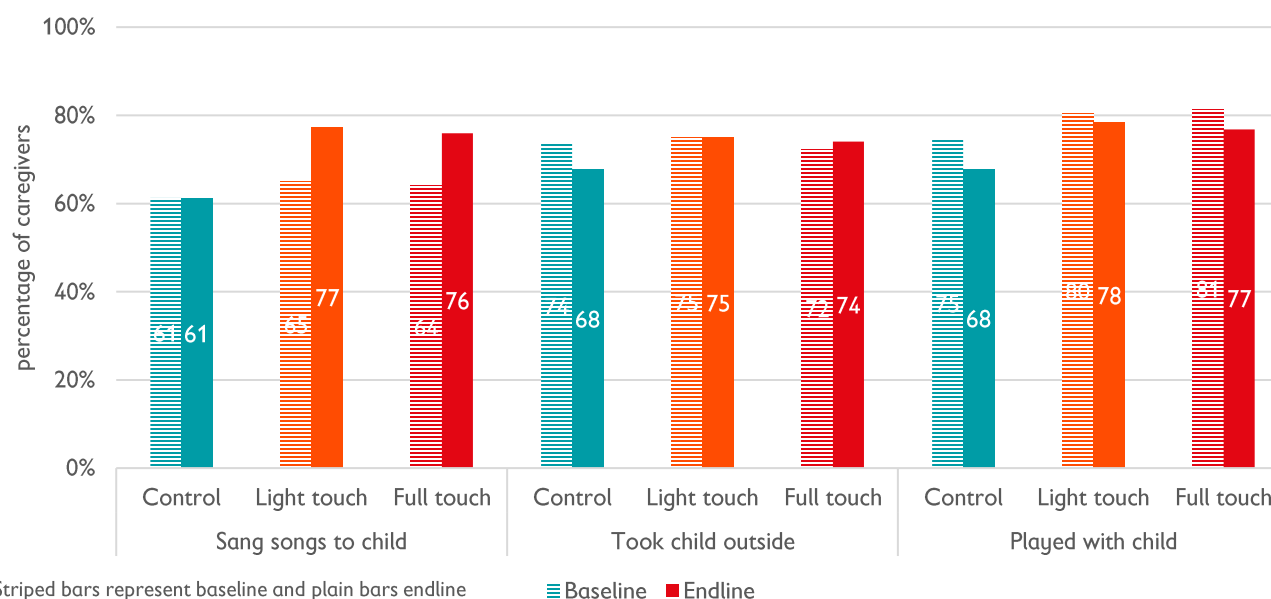


Figure 3: Shares of caregivers reading books, telling stories and naming/counting/drawing with their child, by treatment arm



Note: The difference between the full and light-touch group is not significant.

Figure 4: Shares of caregivers singing songs, going outside and playing with their child, by treatment arm



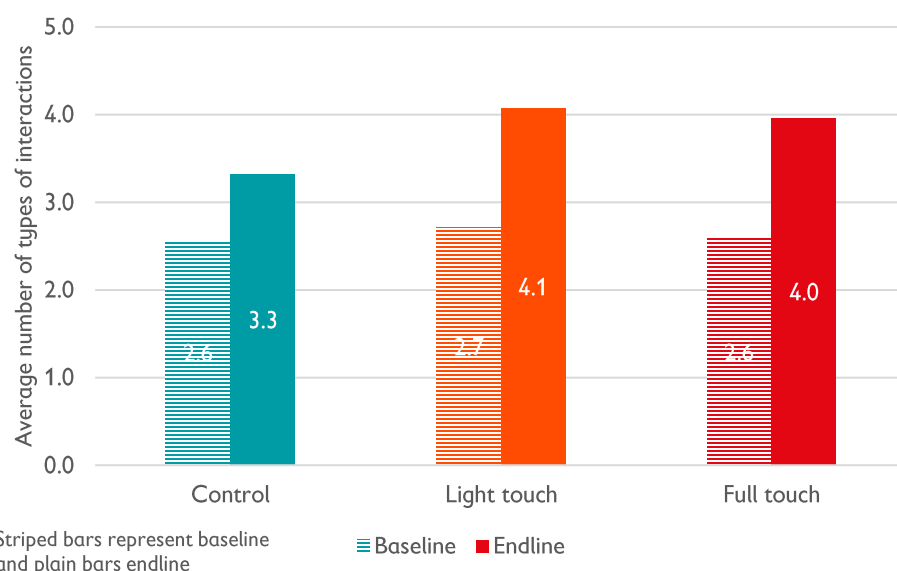
Box 2: Effect of COVID-19 on caregivers' interactions with their child

As a consequence of the pandemic, 62% of caregivers reported increased interest in playing, and 67% reported an increase in time spent interacting with the child. Time spent interacting with the child, in particular, was not comparable across treatment arms: caregivers reported it increased for 70% of households in the two treatment groups (full and light-touch), while it increased for 60% of households in the control group.

The intervention had a positive and significant impact on the involvement of fathers in parenting for both treatment groups. The intervention increased the proportion of fathers involved in at least one activity in the full and light-touch groups, from 33% at baseline to 43% at endline (difference of 10 pp). On the other hand, the same share in the control group did not increase from baseline to endline (37%).

The intervention increased the diversity of types of interactions between the caregiver and child in full and light-touch groups compared to control. The full and light-touch interventions had a positive and significant impact on the total diversity of interactions, calculated as the number of different interactions between any caregiver and the child. The average number of types of interactions with the child increased to 4 in full and light-touch groups, compared to 3.3 in the control group (see Figure 5). In addition, households in either the full or light-touch interventions were more likely to report 5 or 6 interactions of different types at endline, compared to control households. This outcome¹⁹ was included in the analysis to provide a comprehensive summary of the effect of treatment on responsive parenting practices, without focusing on a single activity.

Figure 5: Diversity of total interactions at baseline and endline, by treatment arm



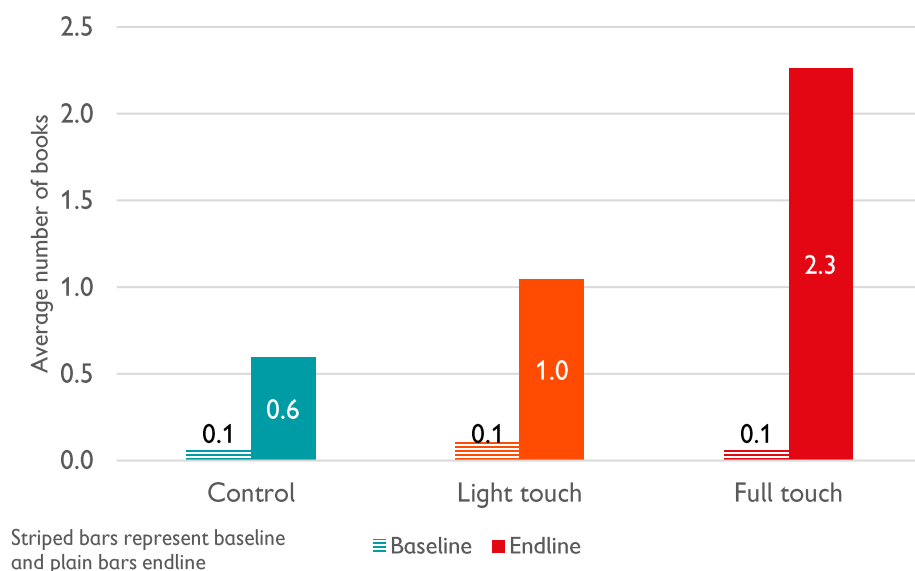
Note: The difference between the full and light-touch group is not significant.

The intervention increased availability of books in treatment households. The availability of books at endline varied drastically across groups: only 27% of households reported their children having books in the control group, while this was 52% in the light-touch group and 71% in the full treatment assignment. The full and light-touch interventions had a positive and significant impact on both the availability of at least one book and the total number of children's books available. The impact for the full intervention was higher than for light-touch - on average at endline households in the full intervention had 2.3 books, compared to 1.0 in light-touch households and less than 1 book in control households (see Figure 6).

Virtually all children in the study at endline played with toys. The most common types of toys were household objects (94%) and homemade toys (90%), while only 35% played with manufactured toys. These shares are comparable across groups.

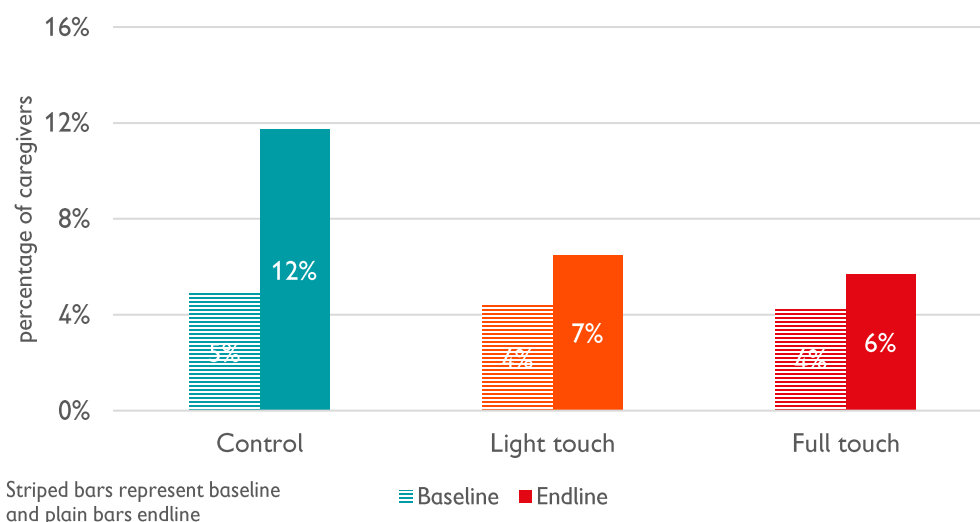
¹⁹ Which was not included in the pre-analysis plan.

Figure 6: Average number of children books per household at baseline and endline, by treatment arm



At endline, relatively fewer caregivers reported leaving the child alone for more than one hour in the treatment groups. 12% of caregivers in the control group reported leaving the child alone, compared to 6% and 7% in the full and light-touch arms, respectively (Figure 7). Leaving the child in the care of another child for more than one hour remains common across treatment groups, with 37% of caregivers reporting doing so in the week before the survey.

Figure 7: Prevalence of children left alone at home for more than one hour at baseline and endline, by treatment arm



Note: The difference between the full and light-touch group is not significant.

Box 3: Effect of COVID-19 on children crying

In the module estimating the impact of COVID-19, 41% of caregivers reported that their child cried more during COVID. However, this was more likely in control groups (47% of caregivers), with lower shares in the groups assigned to the intervention (35% and 40% in the light and full treatment groups respectively).

Attendance to Early Childhood Development (ECD) programs is comparable across groups. At endline, the proportion of children attending ECD programs is similar across groups, at 43% (70% of which are public programs). Programs were rated as generally good (51%) or excellent (34%). There is a significant difference in children attending ECD programs all week at endline: the share was higher in the control group at 73%, compared to 57% in the groups assigned to the intervention (full or light-touch).

2. Positive discipline

The intervention decreased the use of physical punishment in full and light-touch groups but did not have a significant effect on the use of psychological punishment. Both the full and light-touch interventions had a significant and negative impact on caregivers physically punishing²⁰ their children (see Figure 8). The same treatment effect was found for severe physical punishment,²¹ although the share of caregivers using this was only 1% for the control group (and 0% for full intervention and light-touch). On the other hand, the proportion of caregivers using psychological aggression²² decreased for all groups from 31% to 12%, with no impact from assignment to treatment. Attitudes towards physical punishment improved across treatment arms: 21% of caregivers reported they believe a child needs to be physically punished to be raised properly, compared to 36% at baseline.

The use of non-violent discipline in full and light-touch groups increased as a result of the intervention. The full and light-touch interventions had a positive and significant impact on the use of non-violent discipline,²³ whether it was using *only* non-violent methods, or using them among others. At endline, the shares of caregivers using non-violent discipline were 79% in the full intervention group, 84% in the light-touch group and 65% in the control group (see Figure 9).

Caregivers with educational attainment were less likely to improve their positive discipline practices, while caregivers from rural households were more likely to do so. The improvements in the use of physical punishment and the use of non-violent methods resulting from the intervention was significantly lower for caregivers with secondary education compared to caregivers with lower levels of education. On the other hand, caregivers from rural areas experienced a larger increase in non-violent methods, and decrease in the use of psychological violence, compared to caregivers from urban or mixed areas.

Box 4: Effect of COVID-19 on harsh discipline

A decrease in harsh discipline practices has been reported by 52% of households as a consequence of the COVID-19 pandemic. There were no differences in self-reporting across treatment groups. This finding can be reconciled with the national broadcasting of radio programs focusing notably on positive discipline messages during the COVID-19 lockdown period.

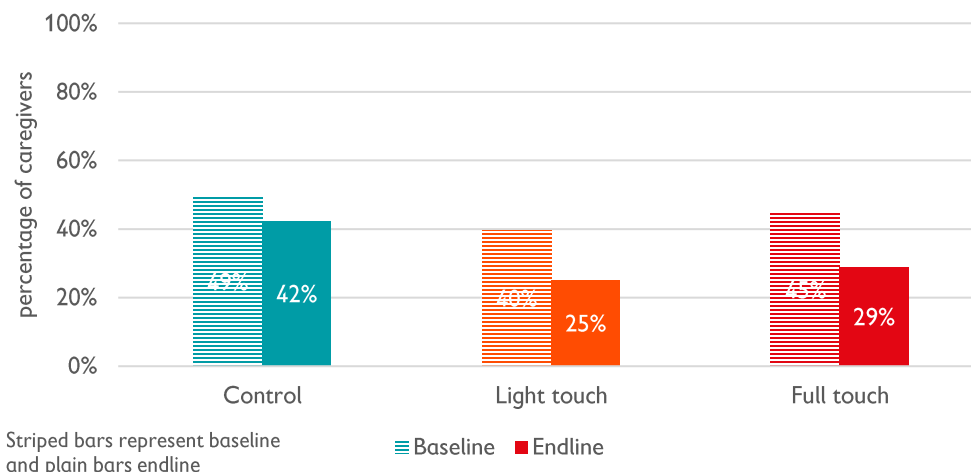
²⁰ Physical punishment includes behavior such as shaking, spanking, hitting, slapping or beating the child when they badly behaved.

²¹ Severe physical punishment only includes hitting on the face, head or ears, or hitting the child repeatedly.

²² Psychological aggression includes shouting/screaming at the child and calling them dumb, lazy or in similar ways.

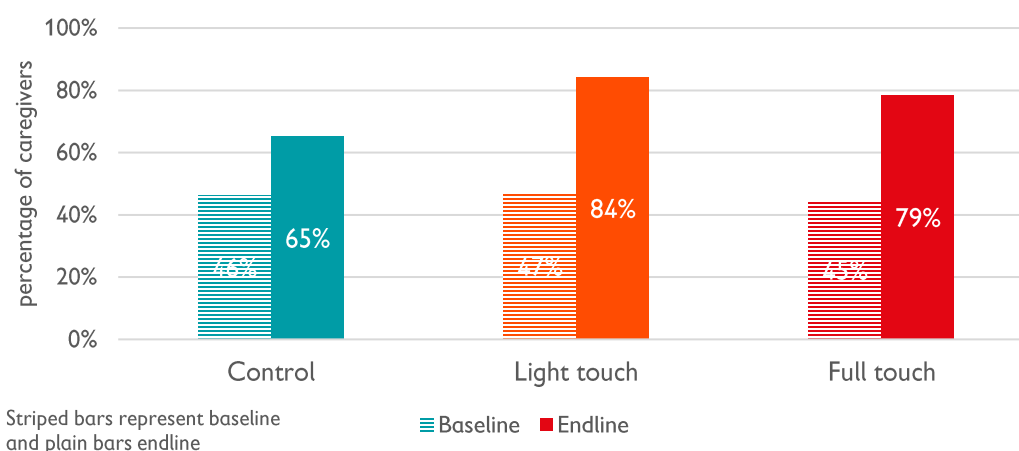
²³ Non-violent punishment includes behavior like taking away privileges, discussing the issue calmly, explaining why their behavior was wrong and giving them something else to do.

Figure 8: Use of physical punishment at baseline and endline, by treatment arm



Note: The difference between the full and light-touch group is not significant.

Figure 9: Use of non-violent discipline at baseline and endline, by treatment arm



Note: The difference between the full and light-touch group is not significant.

3. Health & nutrition practices

Light-touch group caregivers are more likely to meet minimum recommended meal frequency at endline. Children in the light-touch group were significantly more likely to receive three meals per day (the recommended age-appropriate minimum meal frequency) than children from the control group (78% compared to 67% of children in these groups respectively) at endline. The difference for the full treatment group is not significant. Take-up of feeding best practices is likely to have been affected by the COVID-19 pandemic, with 88% of caregivers reporting increased difficulty getting food in this context (see Box 1).

There was no difference in disease prevalence or caregivers seeking treatment for disease across groups. Incidence of diseases was generally comparable across arms, with 15% of children having diarrhea, 24% having a fever, 40% having a cough, 18% having difficulty breathing and 2% experiencing acute respiratory infection in the two weeks before the survey. Proportions of parents seeking treatment for illness was also similar across groups (70% for diarrhoea, 73% for trouble breathing).



Box 5: Effect of COVID-19 on hand-washing

When asked about hand-washing practices at endline, caregivers reported a mean of seven different hand-washing practices in a typical day. The most commonly reported instances of handwashing were after using the toilet (92%), before eating (88%) and after cleaning a child's bottom (88%).

Effects of the First Steps intervention on developmental outcomes

Developmental outcomes are measured through the MDAT, which was administered to a good standard. We normalize each score by child age – using children from the study – so that the outcome is measured in standard deviations from the age-mean. The overall distribution of the MDAT raw and standardized scores is as expected and agreement between assessors observing simultaneously the same child is very high. Details on the quality of administration and of the data are provided in Appendix III. The analysis below excludes outliers, and robustness checks confirmed our findings (see Appendix IV).²⁴ It was performed using both the overall MDAT standardized scores and MDAT sub-domains standardized scores as dependent variables.

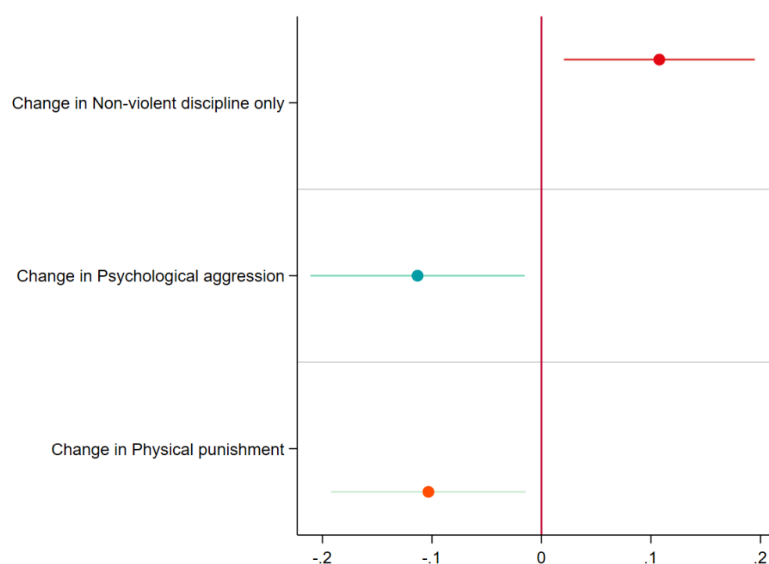
By improving several key caregiver practices, the First Steps program likely contributes to positive developmental changes. We explored which improvements in caregiver practices could be associated with positive changes in developmental trajectories (see Table 11 in Appendix I: Indicator Table) and found that several of them were impacted by the intervention.

First, the intervention increased several responsive parenting practices, which is in turn positively associated with developmental trajectories. Increase in the diversity of interactions caregivers engage in with their child are significantly and positively associated with improvements in child development trajectories. This highlights the importance of increased stimulation at home for strong cognitive development. Since the diversity of interactions increased as a result of the intervention, this result suggests that developmental trajectories will be positively impacted by the First Steps program.

Second, the intervention improved positive discipline practices, which in turn is positively associated with developmental changes. Decreased use of physical violence and psychological aggression and increased exclusive use of non-violent discipline methods are associated with positive development change (see Figure 10). Since the intervention decreased the use of physical violence and increased the use of non-violent discipline methods, this result also highlights a mechanism through which the program will positively impact children's developmental trajectories.

No other change in caregiver practices was found to be significantly associated with change in developmental outcome.

Figure 10: ANCOVA regression coefficients associated with MDAT scores (change in discipline practices)



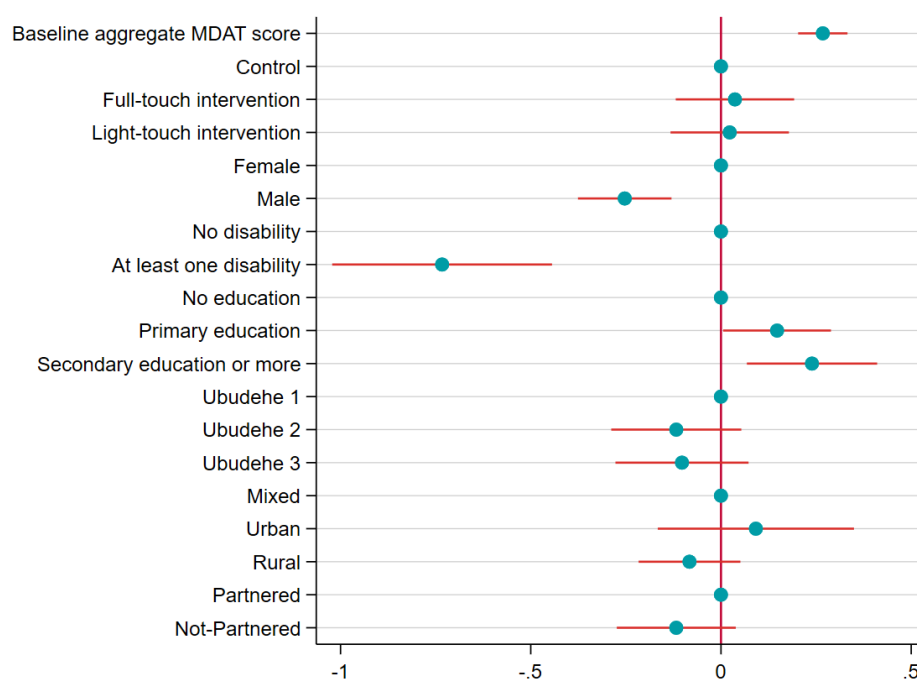
Note: The dots represent the coefficients taken from the ANCOVA model, the horizontal line shows the confidence interval of the estimate. Values for which the estimate and confidence interval fall entirely below zero indicate that the variable is associated with

²⁴ Children whose weighted standardized score is below -3 were removed from the analysis (15 children out of 1,106).

significantly positive change in MDAT scores; conversely, values for which the full confidence interval falls above zero indicate that variable is associated with significantly lower changes in MDAT scores. Among the other control variables included in the model, only the child gender and disability and caregiver completion of secondary education are consistently statistically significantly associated with the endline MDAT score. Coefficients are taken from separate regressions.

Our analysis does not find any significant effect of the intervention on the change in any MDAT domain scores between baseline and endline. Neither the endline MDAT scores nor the change in scores between baseline and endline is predicted by the children's treatment group (see Figure 11 and Table 9 in Appendix I: Indicator Table). This is in line with previous studies assessing the effect of analogous ECD interventions that concluded to a positive impact on parenting KAP but did not detect an impact on the MDAT²⁵. However, this absence of significant effect does not necessarily mean that the intervention does not impact developmental outcomes. First, changes in caregiver practices can take some time to translate into changes in developmental outcomes, and it is thus possible that these changes will materialize in the next few years. Second, it is possible that some developmental outcomes were impacted by the intervention but that the MDAT tool failed to capture it. Jensen SKG et al. 2021 found a significant positive effect of a home-visit based ECD intervention in Rwanda on caregiver practices and developmental outcomes measured through a caregiver-reported tool, but no impact using the MDAT.²⁶ In the absence of anthropometric data, we were not able to explore heterogeneous effects for malnourished children.

Figure 11: ANCOVA regression coefficients associated with MDAT scores (treatment effect)



Note: The dots represent the coefficients taken from the ANCOVA model, the horizontal line shows the confidence interval of the estimate. Values for which the estimate and confidence interval fall entirely below zero indicate that the variable is associated with significantly positive change in MDAT scores; conversely, values for which the full confidence interval falls above zero indicate that variable is associated with significantly lower changes in MDAT scores. Coefficients are taken from the same regression.

Child gender and disability status are strong predictors of change in developmental outcomes. Being a girl is a strong predictor of both higher developmental outcomes (at baseline and endline) and faster developmental trajectories. At

²⁵ In Mozambique, Clacherty et al. 2016 show that the CARE ECD program positively impacts behaviors and reduces caregivers' emotional stress, but no significant effect was detected on cognitive and language development as captured by the MDAT. In Gambia, Blimpo et al. 2019 do not find any significant impact from new community-based early childhood development (ECD) centers, including with intensive provider training, on MDAT scores. In Malawi, Gelli, et al. 2019 do not find any significant impact of adding on MDAT scores, despite significant effects on reported caregiver practices.

²⁶ For a ECD intervention targeted on Ubudehe 1 households in three different districts of Rwanda.

endline, female children score significantly higher at the MDAT than male children: their mean weighted standardized score is 0.3 higher compared to male children (see Table 12 in Appendix I: Indicator Table). Conversely, having a disability is a strong predictor of lower developmental outcomes and slower developmental trajectories, with the exception of socio-emotional skills. The difference in mean standardized scores at endline is about 1 standard deviation. These relationships hold for all developmental domains. The mean MDAT score varies for each type of disability (see Table 13 in Appendix I: Indicator Table), but for all types of caregiver-reported disability children score statistically significantly lower on the MDAT than children with no reported disability.

The educational attainment of caregivers is another strong predictor of change in developmental outcomes, in fine motor skills development in particular (see Figure 11). Children's developmental outcomes at baseline and endline varies with educational attainment of the child's caregiver: the mean weighted standardized score significantly increases by 0.4 for children with caregivers who have secondary education compared to children with caregivers without any education (see Table 14 in Appendix I: Indicator Table). In addition, children whose caregivers completed secondary education developed cognitive skills (and in particular fine motor skills) faster than children whose caregiver have lower educational attainment.

The household geography is associated with higher MDAT scores (see Table 14 in Appendix I: Indicator Table). The average weighted standardized score of children from Gasabo is 0.2 higher compared to children from Ruhango and Kirehe. In addition, children from rural sectors have lower scores than those from mixed or urban sectors. These associations hold more strongly when it comes to fine motor skills and expressive language. However, geographical markers are not significant predictors of *change* in MDAT scores, indicating that their effect was already captured at baseline but did not affect the pace of development afterwards. There are no significant differences in MDAT scores by household Ubudehe category nor caregiver marital status (see Figure 11).

Conclusions

The evaluation confirms that the First steps program positively impacts a range of responsive parenting and positive discipline practices. This evaluation confirms the findings of the pilot study in demonstrating that the intervention leads to significant improvements in caregiver practices. In particular, the intervention increased the variety of interactions caregivers engage in with their child, improved father's involvement in these, and promoted the use of positive discipline methods instead of physical punishment.

Through improvements in key responsive and positive practices, the evaluation indicates a mechanism through which the First Steps program likely contributes to positive developmental changes. Through improvements in these key caregiver practices, the First Steps program likely contributes to positive change in children's developmental trajectories. The diversity of interactions caregivers engage in with their child and the use of positive discipline methods are associated with positive change in developmental outcomes, suggesting that developmental trajectories will be positively impacted by the First Steps program.

The evaluation does not find significantly higher MDAT scores among children of treatment groups. However, this does not necessarily mean that the program did not or will not positively impact developmental outcomes. Both the timing and the choice of the assessment tool determine how accurately the effect of an intervention on children's developmental outcomes is captured.

Overall, the full version of the intervention does not lead to larger improvements than the light-touch version. As there were no systematic differences between treatment arms at baseline, it is unlikely any differences between groups is due to inherent characteristics of the group but instead can be attributed to the impact of the intervention. The evaluation does not find a systematically stronger impact of the full intervention on parenting practices. Where there are differences in the impact on parenting practices between the full and light-touch arms, the differences are not significant. The only exception is the availability of books that was increased to a larger extent by the full version of the intervention, as expected given that the access to a book bank is one of the key differences between the two arms.

Recommendations for management action

The evaluation confirms that the First Steps program positively impacts a range of responsive parenting and positive discipline practices. This key conclusion supports the extension of this program in Rwanda.

The evaluation concludes that the light-touch version of the program could be sufficient to spur changes in practices. The evaluation could not demonstrate a larger impact of the full intervention over the light-touch version (nor the contrary). This suggests that the core components of the intervention present in both versions, the parenting sessions and radio program, lead to the major impacts on caregiver behaviors.

The evaluation highlights opportunities to increase emphasis on a few topics within the program. In particular, while the intervention led to increased use of non-violent discipline methods and decreased use of physical violence, it did not reduce the use of psychological aggressions. This topic is all the more important considering that a decrease in psychological aggression is found to be associated with improved developmental trajectories. We also found limited effects on prevalence of disease and related care-seeking behaviors, suggesting that best practices in these matters could require more emphasis within sensitization sessions.

Future research could explore the medium-term impact of the First Steps program on children's developmental outcomes. Although the evaluation does not detect a significant impact on MDAT scores 18 months after the end of the program, the analysis identifies a mechanism through which the intervention will likely lead to higher developmental outcomes. These effects might take more time to materialize, and future work could investigate the lasting impact of the First Steps program as children continue through their early childhood and into primary school.

References

- Blimpo, Moussa P.; Carneiro, Pedro; Jervis, Pamela; Pugatch, Todd. 2019. Improving Access and Quality in Early Childhood Development Programs : Experimental Evidence from The Gambia. Policy Research Working Paper;No. 8737. World Bank, Washington, DC. © World Bank. <https://openknowledge.worldbank.org/handle/10986/31275> License: CC BY 3.0 IGO
- Boggs D, Milner KM, Chandna J, et al. Arch Dis Child 2019;104:S22–S33. doi:10.1136/archdischild-2018-315431
- Britto PR, Lye SJ, Proulx K, Yousafzai AK, Matthews SG, Vaivada T, Perez-Escamilla R, Rao N, Ip P, Fernald LCH, MacMillan H, Hanson M, Wachs TD, Yao H, Yoshikawa H, Cerezo A, Leckman JF, Bhutta ZA; Early Childhood Development Interventions Review Group, for the Lancet Early Childhood Development Series Steering Committee. Nurturing care: promoting early childhood development. Lancet. 2017 Jan 7;389(10064):91-102. doi: 10.1016/S0140-6736(16)31390-3. Epub 2016 Oct 4. PMID: 27717615
- Clacherty G, Clacherty A, Tredoux C, Dawson M. ECD program impact evaluation report. CARE International, 2016
- Gelli, Aulo and Gladstone, Melissa and Twalibu, Aisha and Nnensa, Theresa and Kariger, Patricia Karol and Alderman, Harold, Adding a Nutrition Behavior Change Communication Component to an Early Childhood Development Intervention in Malawi: A Cluster Randomized Trial (February 7, 2019). IFPRI Discussion Paper 1804 (2019), Available at SSRN: <https://ssrn.com/abstract=3363742>
- Jensen SKG, Placencio-Castro M, Murray SM, et al. Effect of a home-visiting parenting program to promote early childhood development and prevent violence: a cluster-randomized trial in Rwanda. BMJ Global Health 2021 ;6:e003508. doi:10.1136/bmjgh-2020-003508
- Laterite Ltd, Evaluation of Save the Children's Intera za Mbere Program: Baseline Report. Kigali, Rwanda. April 2020
- Leppink, J., O'Sullivan, P., & Winston, K. (2016). Effect size – large, medium, and small. *Perspectives on Medical Education*, 5(6), 347–349. <https://doi.org/10.1007/s40037-016-0308-y>
- Long, Q., Little, R. J., & Lin, X. (2010). Estimating causal effects in trials involving multitreatment arms subject to non-compliance: a Bayesian framework. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(3), 513-531.
- Schochet, P. Z. (2018). Multi-armed RCTs: A design-based framework. *Journal of Educational and Behavioral Statistics*.
- World Health Organization, United Nations Children's Fund, World Bank Group. Nurturing care for early childhood development: a framework for helping children survive and thrive to transform health and human potential. Geneva: World Health Organization; 2018. Licence: CC BY-NC-SA 3.0 IGO.

Appendix I: Indicator Tables

Table 4: Baseline and endline means of responsive parenting practices, by treatment arm

	Control		Light-touch		Full treatment		EL difference across arms
	Baseline	Endline	Baseline	Endline	Baseline	Endline	
Read books to child	21%	53%	21%	64%	19%	62%	*
Told stories to child	12%	32%	8%	47%	7%	44%	***
Sang songs to child	61%	61%	65%	77%	64%	76%	***
Took child outside	74%	68%	75%	75%	72%	74%	
Played with child	75%	68%	80%	78%	81%	77%	*
Named/counted/drew with child	14%	51%	21%	65%	14%	64%	***
Father involved in at least one interaction	37%	37%	32%	45%	33%	42%	
N. of different types of interactions with the father	0.6	0.7	0.6	1.1	0.6	0.8	
N. of different types of interactions	2.6	3.3	2.7	4.1	2.6	4.0	***
Average number of children books	0.1	0.6	0.1	1.0	0.1	2.3	***
Child has book(s)	5%	27%	8%	52%	5%	71%	***
Child playing with toys	94%	100%	95%	100%	93%	100%	
Child left alone at home	5%	12%	4%	7%	4%	6%	*
Child left with another child	33%	43%	34%	36%	26%	33%	

Note: ***, **, and * indicate significance at the 0.1, 1 and 5 percent critical level.

Table 5: Baseline and endline means of positive discipline practices, by treatment arm

	Control		Light-touch		Full treatment		Endline difference across arms
	Baseline	Endline	Baseline	Endline	Baseline	Endline	
Belief in physical punishment	36%	24%	35%	18%	39%	20%	
Physical punishment	49%	42%	40%	25%	45%	29%	***
Severe physical punishment	0%	1%	1%	0%	0%	0%	**
Psychological aggression	30%	12%	32%	11%	34%	14%	
Non-violent discipline	46%	65%	47%	84%	45%	79%	***
Non-violent discipline only	24%	47%	28%	66%	24%	60%	***

Note: ***, **, and * indicate significance at the 0.1, 1 and 5 percent critical level.

Table 6: Baseline and endline means of health and nutrition practices, by treatment arm

	Control		Light-touch		Full treatment		Endline difference across arms
	Baseline	Endline	Baseline	Endline	Baseline	Endline	
Minimum meal frequency	73%	67%	77%	78%	73%	73%	*
Sought treatment for diarrhea	84%	72%	80%	73%	74%	66%	
Sought treatment for difficulty breathing	67%	72%	65%	76%	64%	68%	

Note: ***, **, and * indicate significance at the 0.1, 1 and 5 percent critical level.

Table 7: ANCOVA regression coefficients for full and light-touch groups on responsive parenting practices

	Light-touch		Full treatment	
	Coef.	Std. err.	Coef.	Std. err.
Read books to child	0.543**	(0.18)	0.442*	(0.18)
Told stories to child	0.687***	(0.17)	0.554**	(0.18)
Sang songs to child	0.970***	(0.20)	0.817***	(0.19)
Took child outside	0.388*	(0.19)	0.331	(0.19)
Played with child	0.526**	(0.18)	0.379	(0.21)
Named/counted/drew with child	0.623***	(0.17)	0.596**	(0.19)
Father involved in at least one interaction	0.587**	(0.18)	0.384*	(0.19)
N. of different types of interactions with the father	0.434***	(0.10)	0.216*	(0.10)
N. of different types of interactions	0.682***	(0.13)	0.958***	(0.13)
N. of children books	0.459***	(0.12)	1.692***	(0.26)
Child has book(s)	1.223***	(0.19)	2.121***	(0.21)
Child left alone at home	-0.777**	(0.28)	-0.886**	(0.30)
Child left with another child	-0.235	(0.18)	-0.277	(0.19)

Note: ***, **, and * indicate significance at the 0.1, 1 and 5 percent critical level.

Table 8: ANCOVA regression coefficients for full and light-touch groups on positive discipline practices

	Light-touch		Full treatment	
	Coef.	Std. err.	Coef.	Std. err.
Belief in physical punishment	-0.406	(0.22)	-0.292	(0.23)
Physical punishment	-0.897***	(0.18)	-0.726***	(0.20)
Psychological aggression	-0.369	(0.26)	-0.104	(0.23)
Non-violent discipline	1.225***	(0.21)	0.838***	(0.20)
Non-violent discipline only	0.971***	(0.17)	0.681***	(0.18)

Note: ***, **, and * indicate significance at the 0.1, 1 and 5 percent critical level.

Table 9: ANCOVA regression coefficients for full and light-touch groups on MDAT standardized scores

	Light-touch		Full treatment	
	Coef.	Std. err.	Coef.	Std. err.
age-standardized MDAT score	0.023	(0.08)	0.037	(0.08)
age-standardized Socio-emotional score	0.063	(0.08)	0.125	(0.09)
age-standardized Fine motor score	0.007	(0.07)	-0.015	(0.08)
age-standardized Expressive language score	-0.055	(0.08)	-0.010	(0.08)
age-standardized Gross motor score	0.092	(0.08)	0.015	(0.08)

Note: ***, **, and * indicate significance at the 0.1, 1 and 5 percent critical level.

Table 10: Mean standardized MDAT scores at baseline and endline by treatment group

		Mean	Confidence Interval	
Control	Baseline	-0.03	-0.15	0.09
	Endline	-0.06	-0.17	0.06
Light-touch	Baseline	-0.02	-0.13	0.08
	Endline	0.00	-0.14	0.14
Full treatment	Baseline	0.05	-0.06	0.15
	Endline	0.06	-0.06	0.18

Table 11: ANCOVA regression coefficients associated with change in selected caregiver practices

	Coef.	Std. err.	Confidence interval	
Increase in the number of different types of interactions caregivers engage in	0.035*	(0.02)	0.002	0.067
Increase in the exclusive use of non-violent discipline methods	0.108*	(0.04)	0.020	0.195
Increase in the use of physical punishment	0.103*	(0.04)	-0.192	-0.014
Increase in the use of psychological aggressions	0.113*	(0.05)	-0.211	-0.015

Note: The table shows the coefficient and standard error associated with the change in caregiver practice between baseline and endline. The direction of the coefficient (positive or negative) represents the direction of the effect of that practice on child development. Among the other control variables included in the model, only the child gender and disability and caregiver completion of secondary education are consistently statistically significantly associated with the endline MDAT score.

Table 12: Mean endline standardized MDAT score by child characteristic

	Mean	Confidence Interval	
Gender			
Female	0.19	0.10	0.29
Male	-0.16	-0.26	-0.07
Disability status			
No disability	0.06	-0.01	0.13
At least one disability	-0.90	-1.18	-0.63

Table 13: Mean endline standardized MDAT scores by caregiver-reported disability

	Mean	95% CI	
Serious delay in sitting, standing, or walking	-0.75	-1.16	-0.33
Difficulty seeing, either in the daytime or at night	-0.42	-0.86	0.02
Have difficulty hearing (uses hearing aid, hears with difficulty, or completely deaf)	-0.17	-1.04	0.70
Have difficulty in moving his/her arms or does he/she have weakness and/or stiffness in the arms or legs	-0.74	-1.42	-0.06
Sometimes have fits, become rigid, or lose consciousness	-0.34	-0.77	-0.10
(Does not) learn to do things like other children his/her age	-0.63	-1.37	-0.11
Appear in any way mentally dull or slow (other cognitive disability)	-1.18	-1.67	-0.68
At least one disability	-0.76	-1.01	-0.50

Table 14: Mean standardized MDAT score by household characteristic

	Mean	Confidence Interval	
Education			
No education	-0.19	-0.33	-0.06
Primary education	0.01	-0.09	0.10
Secondary education or more	0.23	0.11	0.36
District			
Gasabo	0.14	0.04	0.25
Ruhango	-0.09	-0.20	0.03
Kirehe	-0.08	-0.21	0.04
Sector urbanization status			
Rural	-0.09	-0.19	0.02
Mixed	0.07	-0.03	0.17
Urban	0.17	-0.09	0.44

Table 15: Household characteristics of the whole endline sample and households with a change in caregiver

	Full sample	New caregiver
Ubudehe		
1	13%	9%
2	53%	64%
3	33%	27%
Caregiver education		
No school	30%	30%
Primary	48%	45%
Secondary and more	22%	25%
District		
Gasabo	38%	39%
Kirehe	39%	45%
Ruhango	23%	17%
Marital status		
Partnered	80%	56%
Non-Partnered	20%	44%

Appendix II: Sample Size Calculations

Assuming a two-sided t-test with a 5% significance level, 80% power, and intra-cluster correlation (ICC) of 0.1, this research design is powered to detect a difference greater than 0.289 standard deviations between any two arms of the trial. Based on the magnitude of the coefficients of the impact of the program on Ages & Stages Questionnaires (ASQ) scores available from the pilot evaluation, a minimum detectable effect (MDE) smaller than 0.3 standard deviations is sufficient. We do not have sufficient power to report results by district and only report overall results for the study.

The variance of the estimated treatment effect can be calculated with the following formula:

$$SE^2 = \frac{4\rho\sigma_y^2}{K} + \frac{4(1-\rho)\sigma_y^2}{nK} = [(n-1)\rho + 1] * \frac{4\rho\sigma_y^2}{nK}$$

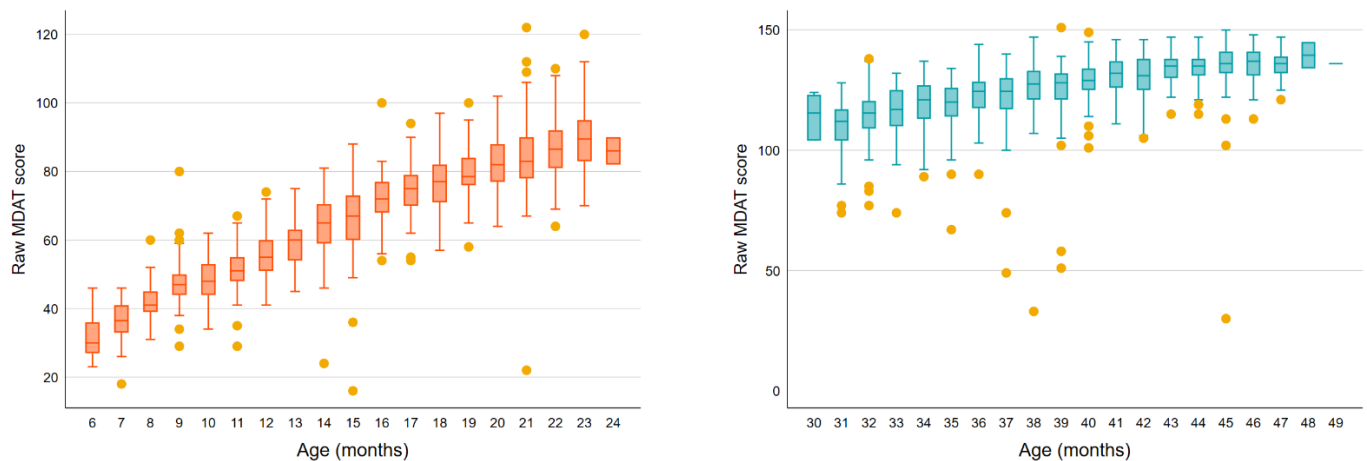
where σ_y^2 is the total outcome variance within each treatment arm, nK is the total sample size of the trial, and ρ is the ICC. The first half of the equation shows that the total variance of the treatment effect is the sum of sampling error between clusters and that within clusters. The factor $[(n-1)\rho + 1]$ in the last half of the equation is the design effect (DE). A smaller variance of the treatment effect means that the estimation is more precise and the analysis can detect a smaller MDE (or is more powered to detect a given MDE).

Appendix III: Quality of the MDAT administration

To assess the quality of administration and of the data, we looked at the following factors.

The MDAT scores increase with age, as is expected for this assessment. Raw scores (the number of items a child was able to perform) on the MDAT are age-dependent, as expected for this child development assessment (see Figure 12 & Figure 13). This trend can also be seen within each domain of the MDAT, although the slope is steeper for fine motor and expressive language skills.

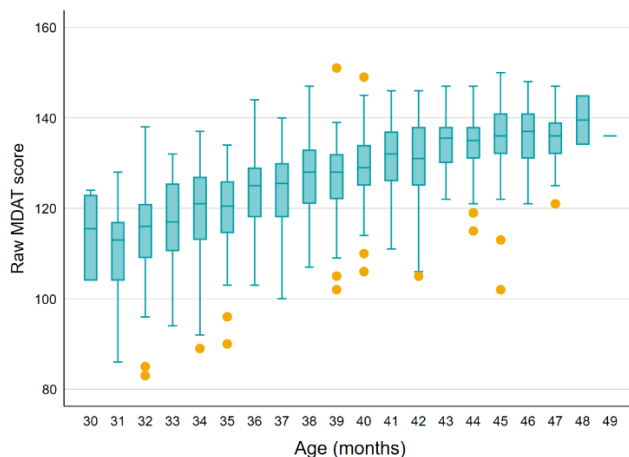
Figure 12 and Figure 13: Distribution of the baseline (left) and endline (right) raw MDAT scores by age



Note: The median and distribution of MDAT raw total scores increases with age of the child. The middle bar is the median, the shaded box represents the middle 50% of the data. The whiskers represent the limits of the range defined by 1.5 times the inter-quartile range, and the dots denote values outside this range.

Outliers are excluded from our core analysis. There are some outliers in the data, for example cases where the child was not able to perform certain tasks due to a disability, or where the child was irritable or tired and refused to participate in certain tasks. We also present the re-computed endline distribution of standardized scores excluding outliers (see Figure 14), which we define as standardized scores below -3.

Figure 14: Distribution of the endline raw MDAT scores by age, excluding outliers



Given that MDAT scores are age-dependent, scores are age-standardized for the analysis. To derive age-standardized scores, we calculated the weighted mean and weighted standard deviation of scores within each one-month age category. From these statistics, we calculated a weighted Z-score for each child. This process of standardizing scores allows us to have a comparable score across age groups, which is beneficial for comparing associations between variables in the dataset. However, it can be difficult to translate standardized scores or differences in standardized scores between groups into a meaningful indicator, such as months of development or percentile score, as each child's score is relative to those in their age group. Nevertheless, one convention used in social sciences is that a difference in scores of 0.2 is considered a small effect size; a difference of 0.5 a medium effect size; and a difference of 0.8 a large effect size (Leppink, O'Sullivan, & Winston, 2016).

MDAT scores have an approximately normal-shaped distribution, which is as expected for this instrument. Similarly, for each domain we see approximately normal distributions of the (weighted) standardized scores. The presence of outliers is reflected in the left tail of the distributions, which are limited when excluding children with disabilities. In our analysis, we excluded children with a standardized weighted score below -3 before recomputing age-standardized MDAT scores.

Figure 15 & Figure 16: Distribution of baseline (left) and endline (right) weighted standardized MDAT scores

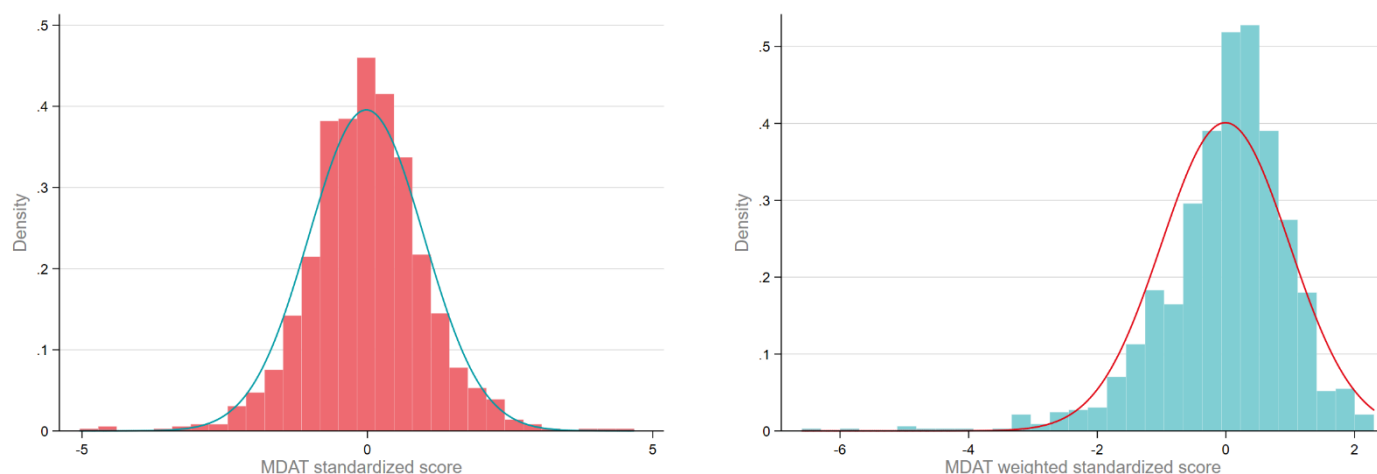
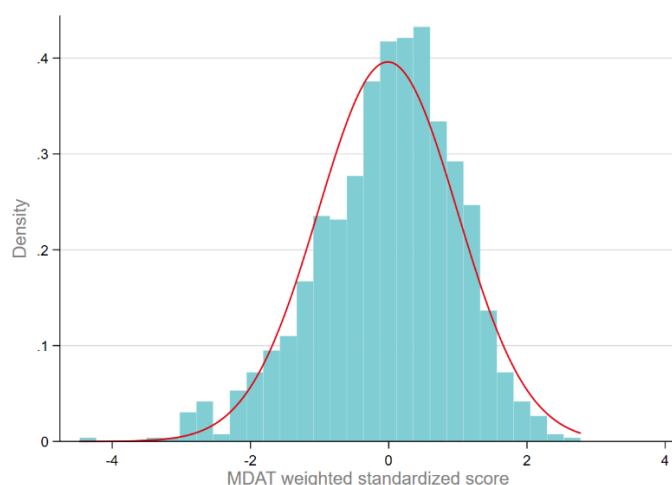


Figure 17: Distribution of endline weighted standardized MDAT scores, excluding outliers



Enumerators were consistent in administering the MDAT. Because the MDAT is a complex observational assessment, we assessed consistency in administration between enumerators during pilot, and the first week of data collection to

provide a measure of confidence in the quality of the assessment administration and, where inconsistent, to provide feedback to the enumerators. Enumerators conducted paired interviews, in which one enumerator conducted the assessment as usual and another silently observed and independently recorded if the child passed or failed each item. To ensure unbiased results, enumerators were seated in a way that would not allow them to look at their partner's responses. From this paired data, we calculated percentage agreement and reliability, measured using Kappa, between paired enumerators. The results by assessment domain for pilot and data collection are shown in Table 17 and 18. Overall, the percentage agreements were very high and the Kappa values indicate near perfect agreement.²⁷ These results show high consistency in administration of the instrument and are a marker of high data quality.

Table 16: Consistency in enumerator assessments on the MDAT (pilot)

domain	agreement	expected agreement	kappa	std_err	z_stat	N
Socio-emotional	99%	67%	96%	0.04	22.7	564
Fine motor	99%	55%	97%	0.04	24.5	637
Expressive language	99%	62%	98%	0.04	25.7	692
Gross motor	97%	66%	92%	0.04	20.6	495

Table 17: Consistency in enumerator assessments on the MDAT (data collection)

domain	agreement	expected agreement	kappa	std_err	z_stat	N
Socio-emotional	97%	59%	93%	0.05	20.4	488
Fine motor	97%	54%	95%	0.04	21.4	511
Expressive language	99%	59%	97%	0.04	23.4	575
Gross motor	99%	62%	97%	0.05	20.0	427

²⁷ We also calculated percentage agreement by item and by enumerator, that were all very high.

Appendix IV: Robustness Checks

Analysis of the Average Treatment effect on the Treated (ATT)

We use an Average Treatment effect on the Treated (ATT) analysis to estimate the results of the intervention on respondents who reported receiving the treatment. Our main estimator, the ITT, is the best-suited to inform a possible scaling-up of the intervention, as it includes all individuals initially allocated to each group, regardless of whether they received/attended the intervention. However, since non-compliance is substantial (see more details in the Limitations section), the ATT can deliver additional information on the intervention effects considering actual implementation and take-up. The ATT estimator compares the outcomes of households that were actually exposed to the program²⁸ with the outcomes of households that did not, regardless of their study group assignment.

Since the ATT analysis relies on self-reported information and is based on self-selection, results should be interpreted in the context of its limitations. The ATT estimator cannot be generalized to the population of interest because of selection bias: since treatment assignment is based on actual attendance to parenting sessions, caregivers in the full or light-touch groups will be those that chose to attend. This makes them systematically different from caregivers in the control group. It also implies that estimates of the impact of the treatment obtained through the ATT analysis will be larger than those obtained through ITT, since individuals in the full and light-touch treatment groups are likely to be more motivated than those in the control group. Furthermore, this ATT analysis relies on self-reported attendance to parenting sessions and home visits that occurred more than a year and a half earlier, and so categorization to treatment group for the purposes of this analysis is highly susceptible to memory bias.

As expected, the ATT analysis confirms that the intervention had an impact on various caregiver practices, but finds a larger effect of the full intervention than for the light-touch version. The full and light-touch interventions increase the share of caregivers interacting with their child in all domains, with the impact of full treatment assignment being consistently higher than the one for light-touch. In addition, the ATT analysis finds that the intervention decreased belief in physical punishment for the full and light-touch groups, while no impact was found with the ITT analysis. These results are expected and at least partly explained by self-selection into the program and memory bias: on the one hand, caregivers who are more likely to improve their practices are more likely to be exposed to the intervention (treatment caregivers complying with their assignment or contaminated control caregivers); on the other hand, caregivers who benefit the most from the intervention are more likely to recall the program. Nonetheless, the larger impact found for the full intervention could also mean that our ITT analysis could have detected significantly larger effects from the full intervention, had compliance been higher.

The ATT analysis confirms the absence of effect on MDAT scores, with the exception of socio-emotional skills. Using self-reported exposure to the intervention, we did not find any significant effect of treatment on the overall standardized MDAT score. The same holds for the different domains, except for socio-emotional skills, for which we found an effect of the full intervention, significant at the 5% level. The limitations described above, including memory and self-selection biases, are similar for this analysis.

Analysis of the Local Average Treatment Effect (LATE)

The Local Average Treatment Effect (LATE) analytical approach does not display the ATT self-selection flaws. The LATE can be conceptualized as the effect of the intervention on compliers: households that were assigned to the full intervention and attended parenting sessions and received a visit, or households that were assigned to the light-touch group and attended parenting sessions. The LATE estimator takes into consideration original treatment assignment and it

²⁸ In this framework, the households that received the full treatment are those who reported attending at least one parenting session and who reported receiving at least one visit from IZU. Households in the light touch group are those who reported attending parenting sessions and who reported not receiving any visit from IZU.

corrects for the self-selection bias present in the ATT. However, since it only estimates treatment effects for a sub-sample (compliers), the results of this analysis also cannot be generalized to the population of interest.

Because of the substantial crossovers across treatment groups, the LATE could not be estimated in this study. Estimating the LATE is usually straightforward in single-arm RCTs (with only one treatment version), where non-compliance only refers to treatment participants not receiving treatment or control participants receiving treatment. However, more assumptions are needed when dealing with multi-arm RCTs (multiple treatment groups). The most important assumption has to do with the absence of *crossovers*, i.e., households that were assigned to one treatment group but actually receive another version of the intervention²⁹. In the case of this study, self-reported non-compliance (outlined in the Limitations section) indicates that the assumption is not met. Indeed, 20% of caregivers in the light-touch group reported receiving the home visit component of the full intervention, and 43% of caregivers in the full treatment group reported not receiving the home visit component that differentiates full intervention from the light-touch.

Heterogeneous Treatment Effects

We investigate the heterogeneous impact of the intervention to understand if some sub-groups benefited more (or less) from treatment assignment. We analyzed the effect of the intervention by district, sector urbanization status, household socio-economic status and child disability status. To eliminate the risk of endogeneity, we defined all these indicators using baseline data only. However, it should be noted that the study was not adequately powered to detect small effects within sub-groups.

The impact of the intervention on caregiver and children's outcomes is generally the same for all households in the sample. Due to the high number of outcomes in the caregiver section³⁰, we look at heterogeneous treatment effects for only four main outcomes: use of non-violent discipline methods, number of children books, number of total interactions with the child, and leaving the child alone for more than one hour. There is no sub-group for which the intervention had a smaller or larger impact in all of those outcomes. When looking at the increase in the number of books, the increase in the full intervention group is larger in the Kirehe and Ruhango districts compared to the Gasabo district, while the increase in the light-touch group is significantly larger in the Kirehe district compared to other districts. The heterogeneous treatment effect analysis confirms the absence of effect of the treatment on MDAT scores.

Outliers in the MDAT distribution

The evaluation findings are robust to inclusion of outliers and alternative definitions of outliers on the MDAT. These outliers are, for example, cases where the child was irritable or tired and refused to participate in certain tasks. Our main analysis excludes children whose standardized score is below -3 at endline. The distribution of endline raw and re-computed standardized scores excluding outliers can be found in Figure 11 and 14 in Appendix III. To check the effect of these outliers on the analysis results, we performed our core analysis on several samples, including all children, removing children with reported disabilities, or excluding both children scoring below -3 and children with reported disabilities. The main result - an absence of effect of treatment on MDAT scores - was consistently confirmed for each sample.

²⁹ In this case, crossovers are defined as (1) households that were assigned to the full intervention, but only attended parenting sessions, and (2) households that were assigned to the light-touch intervention, but also received home visits.

³⁰ The number of caregiver outcomes in the analysis is 20, which, combined with the 4 sub-groups of interest, would amount to 80 regression analyses for heterogeneous treatment effects. Such a high number of outcomes would create the risk of finding significant results purely by chance.

Appendix V: Terms of Reference

Laterite Ltd. was contracted by Save the Children International (SCI) in 2019 to conduct this study, which included:

- designing the study, including sampling and instruments, with close review from SCI;
- securing research approvals, in collaboration with SCI;
- leading field activities, including enumerator training, field preparation, pilot and data collection activities;
- leading the data analysis and reporting of this study, with close review and validation from SCI.

Appendix VI: Data Collection Instruments

Data collection instruments administered at endline are attached to this report.

