



Save the Children®

**Reducing the effects of toxic stress in young children
on the move through positive parenting and resilience
building in Iraq**

Endline Report

December 2021

Mohammed Eldesouky & Allyson Krupar

Reducing the effects of toxic stress in young children on the move through positive parenting and resilience building in Iraq

Executive Summary

The “Reducing the effects of toxic stress in young children on the move through positive parenting and resilience building in Iraq and Colombia” program was launched in December 2020. The primary focus of the program was to build resilience skills in young children, promote positive parenting and disciplines across primary caregivers, and transform male caregiver rearing practices and gender norms around participation in the early years. The program implemented activities in two country contexts:

- Iraq: Duhok refugee camps
- Colombia: Six informal settlements in Montecarlo, Villa Madre Laura, Torres de la Majayura, La Pista, Bendicion de Dios 1, and Bendicion de Dios 2) in Maicao, La Guajir

This report focuses on data collected in Iraq using household based face-to-face assessments following COVID-19 protocols.

The objectives of this study were to measure change in:

- Caregiver participants’ at home practices, positive parenting practices, play-based interactions, self-care and regulation over the course of the intervention
- Male caregivers behavior that fosters non-violent, positive home environment
- Children’s learning and development outcomes measured by CREDI (0-3) and IDELA (3-5)

The goal of the objectives is to understand the relationship between changes in caregivers’ practices and children’s early learning and development.

This study used a longitudinal mixed-methods approach to document changes in perceptions and practices around early child development and caregivers’ practices in communities implementing the ECD Toxic Stress Mitigation Toolkit. The study focused on caregivers and parents of young children, and children aged zero to six years old.

Iraq Sample:

The baseline sample in Iraq include 495 caregivers and 218 3 to 6 year old children children. At endline, 302 caregivers were matched

The study results displayed a significant and positive correlation (effect size) with four caregiver outcome indicators: caregiver child relationship, learning and playing, caregiver resilience, caregiver self-efficacy. On the other hand, the findings were insignificant for the remaining three outcomes: gender-based attitudes, gender-based practices and caregiver perceived stress, suggesting no program effect on these three outcomes.

In addition, the heterogenous effect estimates, exploration of effect size by gender, showed that male caregivers had a larger and statistically significant effect size for: caregiver-child relationship, caregiver resilience and caregiver self-efficacy. Female caregivers displayed a higher effect size on only one

outcome-- learning and playing activities. Moreover, educational level was a significant predictor of the seven outcomes. family size was, as well, a strong predictor of caregiver perceived stress.

The above results were consistent with the findings from the qualitative findings where respondents reported an improved family dynamics and reduced violent child rearing behaviors. They also mentioned that the program was particularly beneficial for male caregivers as they became more aware of the importance of their involvement in child development and became more engaged in it. This suggests changes toward more gender equitable practices

Although, this study used a quasi-experimental design to make causal inferences, we had several limitations related to sample size, omitted variable and model assumptions that limits interpretation of the findings to correlations rather than a straightforward causality. With those limitations in mind, causal inferences about impact of the program on children reveal that:

- Children between 0-3 years old experienced no significant effect size from program participation on their outcomes; neither on the main CREDI score or domains' scores.
 - Analysis of child development predictors for children between 0-3 showed that gender-based attitudes is a predictor of CREDI's overall and domain scores.
 - Testing for mean difference of CREDI scores by gender provides no evidence for difference between female and male children's average scores.
 - Children between 4-6 years old experienced a positive and statistically significant correlation between the intervention and children's overall IDELA score.
 - Repeating the analysis to IDELA's domains scores, we concluded a statistically significant positive relation between the program and two outcomes: emergent numeracy and social-emotional. On the other hand, the changes occurred on the remaining four outcomes could not strictly be attributed to program participation since these changes lack statistical significance.
- Heterogenous effect analysis for IDELA scores showed that female children had higher effect size than male children whose effect size was not significant. In a similar way, comparing the effect sizes between female and male caregivers shows that children of female participants displayed on average a higher and statistically significant (5%) effect size than children of male participants, whose results were not statistically significant.

Quantitative and qualitative data suggest that the program has made contributions to behavior change for caregivers of young children in Iraq. They also displayed a better caregiver resilience scores. Caregivers and parents did not show more equitable gender attitudes or better stress management scores, but they did register more equitable gender practices by the end of the ECD program.

Behavior change takes time and reinforcement, while we see improvements in caregivers' practices and the home learning environments, these findings suggest that perhaps some messages about some responsive care and gender practices were not yet internalized by caregivers. In addition, the absence of program effect on parents' stress may have been impacted by the fact that families are experiencing strong financial stresses. Continued research and programming will be necessary to achieve the desired behavior change, including the design of activities/platforms that might increase engagement among male caregivers.

Introduction

The Early Childhood Development (ECD) program in Iraq launched in December 2020. The primary focus of the program was to build resilience skills in young children, promote positive parenting and disciplines across primary caregivers, and transform male caregiver rearing practices and gender norms around participation in the early years. The program implemented activities in two refugee settlements in Duhok utilizing face-to-face and remote outreach to as many families as possible given the COVID-19 restrictions. Data was collected face-to-face.

This project aimed to provide opportunities to build resilience and mitigate the negative developmental effects of adversity that can stem from forced displacement. Specifically, this project supported populations on the move in Iraq by introducing a home-based approach that promotes ECD as a protective mechanism, to support young children's social-emotional skills and wellbeing.

[Structure of the program-briefly]

The objectives of this study were to 1) monitor changes in child development outcomes; 2) monitor changes in caregivers' attitudes and behaviors who were engaged in the ECD program; and 3) understand the relationship between changes in caregivers' practices might influence children's early learning and development.

Methods

This study used a longitudinal mixed-methods approach to document changes in perceptions and practices around early child development and caregivers' practices in communities implementing the ECD Toxic Stress Mitigation Toolkit. The study focused on caregivers and parents of young children, and children aged zero to six years old. Tools used and samples obtained are described in detail in the following sections. A pre- and post- quasi-experimental design with a comparison group was used to measure changes in children's development and caregiver behavior across the two periods and between the comparison and intervention groups. The difference in differences (DID) technique was used to analyze the quasi-experiment design to approximate the effect size of program participation on child level outcomes and caregiver practices. The supplementary qualitative research was conducted with caregivers and parents to triangulate information and inform program successes and areas that needed to improve to achieve immediate and lasting changes.

Tools

Child early learning and development

For children who were 0-36 months of age, we administered the Caregiver Reported Early Development Index (CREDI) tool is used to measure child development and learning in young children. CREDI is a child development instrument developed by Harvard University's Center for the Developing Child. The CREDI tool contains 146 questions in four domains: motor, language, cognitive and social-emotional development.

The CREDI tool was adapted to the Colombia context and translated sensitively to ensure that the meaning of the questions remains the same across cultures and contexts.¹

For children who were 4 years or older, we administered Save the Children's International Development and Early Learning Assessment (IDELA). IDELA includes 5 SEL items, 7 emergent numeracy items, 6 emergent literacy items, 4 motor skills items. IDELA total score was calculated by adding the weighted total scores from the core domains (motor, literacy, numeracy, and social-emotional) so that all domains contribute equally to the total score.

Caregiver questionnaire

A multifaceted questionnaire was used to gather information about caregivers' practices with their children and other family characteristics (Table 1).

Table 1. Caregiver questionnaire

Section (s)	Description
1. General family information	Parental age, parental literacy, parental education, disability, number of children at home, housing materials
2. Home learning environment	Types of reading materials at home, types of toys at home
3. Early stimulation, responsive care, and resilience	Adults in the home engaging with children to promote early stimulation
4. Parenting self-efficacy ²	Self-efficacy in specific parenting tasks (e.g., soothing the baby; feeding, changing, and bathing the baby)
5. Perceived stress scale ³	Degree to which situations in your life are perceived as stressful
6. Gender attitudes and practices	Attitudes and practices related to decision making, care of children, and involvement in children's lives

Related to the parenting self-efficacy and gender-specific scales, first participants were asked whether they strongly agree, agree, disagree, or strongly disagree with a series of gender attitudes. Additionally, parents and caregivers responded how frequently (never, almost never, sometimes, fairly often, very often) they were engaging in conversations and interactions with their partners about children and their care. For this analysis, the responses were coded on a scale of 0 to 4, a number closer to 4 implies either more gender-equitable attitudes/practices or more self-efficacy in specific parenting tasks.

¹ The CREDI scoring algorithm used for this report reflects the CREDI team's most recent recommendations. In contrast to previous reports (which used a much simpler scoring method), the CREDI scores discussed in this report are more robust and reflect the fact that individual skills can indicate developmental progress in more than one domain of development. For example, a question about whether a child can draw with a stick or pen reflects both their motor and cognitive development. The new scoring algorithm allows us to quantify these contributions. Midline and baseline scores used in this report were re-processed according to these new methods.

² Parenting self-efficacy. Teti D, Gelfand D. Behavioral competence among mothers of infants in the first year: The mediational role of maternal self-efficacy. *Child Development*. 1991;62:918–929

³ Perceived stress scale. Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 24, 385-396.

Structured Interviews

Interviews were used to understand caregivers' perceptions of the benefits and challenges of this new program in more depth. Interviews included approximately 8 discussion questions related to the perceived value of the initiative and perceived changes in caregiving practices.

Sample

Households were randomly drawn from two camps (Gawilan and Bardash) near Duhok to reach a representative sample of the total 500 households reached in program activities. An index child was identified in each household ranging in age from 0 to 5 years old. In households where more than one child is in this age range, enumerators selected the one child ages 0-3 and one child ages 3-5 to interview. Male and female caregivers were interviewed in each household with 117 households having 2 caregivers interviewed, 141 households with only male caregivers interviewed, and 121 households with only female caregivers interviewed. The original sample aimed to have only one caregiver in a household assessed, but due to under involvement of male caregivers without the female counterparts, the program implementers included both male and female caregivers in data collection and program activities.

To calculate the sample size for the quasi-experimental design, we conducted power calculations using Optimal Design software. We assume a $R^2=0.3$. With this assumption and for a 5% significance level ($\alpha = 0.05$) and 80% power ($\beta = 0.8$), the Optimal Design software suggests that we will require at least 247 (for both intervention and comparison groups) children to detect an effect size of 0.3 SDs. 125 children were identified in ages 0-3 and 125 children were identified in ages 3-5, with equal representation in both groups of boys and girls. The comparison group included the same demographic structure (250 children, 125 in both age groups, and equal representation of boys and girls). Given the small samples of boys/girls by age in the final sample, power will be reduced in these disaggregated analyses.

For the qualitative research, we sampled 14 caregivers using a positive and negative deviance approach. Households were selected based on the average value of caregiver-reported practices at baseline. 21 caregivers participated in interviews with observation at baseline but only 14 were tracked to endline.

Prior to the data collection, enumerators attended a two-day training for quantitative data collection. The training was on how to administer the tools. The training consisted of one day reviewing the tools and one day practicing with the tools by conducting mock interviews. Data was collected in households in person following COVID-protocols.

Ethical considerations

The study was submitted to the Save the Children US Ethics Review Committee, the assigned reviewer granted final approval in November 2020 (SCUS-ERC-FY2020-118) for data collection in Iraq. Caregiver and then child verbal assent was requested prior to starting the survey questionnaire.

Limitations & Model assumptions

- a- The Diff-in-Diff method is parametric, meaning that there are set of assumptions that must be relaxed to guarantee the internal validity of the model (generating unbiased effect estimates). Three key assumptions are parallel trend, common shocks and no-spill over.

- Parallel trends, although it is a counterfactual assumption and could not be tested, there are some quantitative methods that could be used to provide some evidence for. Unfortunately, these methods require more than two periods of data which was not the case in our data (only two periods). This affects the study in a way that any estimation of effect size must be cautiously interpreted and reported (e.g. not interpreted as a casual relation, but rather as a possible correlation).
- Common shocks, is a supplementary assumption to parallel trends, and states that exogenous shocks affect both comparison and intervention groups equally. Relaxing this assumption provides additional validity to parallel trends. However, we could not test it empirically as it requires the presence of both historical data and exogenous shocks.
- No-spill over, this assumption is challenging to test as intervention and comparison groups come from the same two camps, and there is no evidence that program people don't have connection with those in the comparison, and could communicate the treatment's effect to them. Additionally, there was not any observed variable in our data that could be used to measure spillover magnitude quantitatively. Accordingly, we can neither confirm nor reject the presence of spillovers in our study.

b- Omitted variable bias (OVB) is an intrinsic limitation of quasi-experiment methods (where no random assignment), which stresses that changes in outcomes could be attributed to unobserved characteristics/variables in the data. We have no reason to believe that our estimating models don't suffer from OVB, particularly with variables that change with time (time-variant). For instance, at best, our model explains 17% of variations in outcomes, leaving more than 80% of outcome variation unexplained by our data. OVB is another reason to interpret diff-in-diff estimations as correlations rather than causation. For future studies, data collection efforts are recommended to gather information on background characteristics that show strong relevancy to outcome variables, aiming to minimize OVB as possible.

c- The study's sample was drawn at random, which conclude to that the two camps' population are expected to exhibit similar outcome patterns as our sample, but not exactly the same. This is because of two reasons, first, the two camps were sampled disproportionately (one camp was represented more in the sample than the other), which was not warranted by actual population proportions in real life. Secondly, research architect deliberately designed the experiment to have equal male to female children proportions, which was not warranted by actual population proportions. In addition, within the analysis, the disproportionate sample was not accounted for using any sample weights. All combined affect the randomness of the sample, therefore, the representativeness of the study sample.

d- Attrition rate with displaced population is expected to exist to a certain extent, however, its magnitude is unknown until it happens. within this study, the overall attrition rate was 24% (nearly a quarter of the sample was not observed in the endline data). Part of this attrition was due to the fact that two caregivers belonging to the same household were surveyed instead of surveying identical observations. This created redundancies, which we had to omit at some point for the analysis. Attrition impacts the analysis in two ways: biasing the effect size estimates if the attrition was selective (based on certain background characteristics), and reduces statistical power. Our investigation to attrition provided no evidences that attrition was selective, thus does not bias our

estimates. On the other hand, the study suffered a loss of statistical power. Future studies can account for attrition pre-intervention by inflating the sample size acquired through the power calculation with the attrition rate found in this study (24%).

In light of these limitations, it is inevitable to interpret the effect size estimates laid out within this report as correlations (association/relationship between program participation and changes in outcomes) instead of straightforward causality (program participation would definitively affect participants' outcomes in a defined and specific manner).

Attrition

Aligning with the study design, we analyzed two types of attrition; **overall attrition** which is the total loss of data between the two periods, and the **differential attrition** which refers to the difference in the rate of attrition between intervention and comparison groups. Since selective attrition could bias effect size estimates, the purpose of computing the two attrition types is to determine whether the attrition rate in our data is within an acceptable range. The acceptable range of attrition is introduced by HomVEE attrition standards, established by the U.S. Department of Education's Institute of Education Sciences (2015). We estimated the overall attrition to be 24% (split 50% to 50% between intervention and comparison groups). The differential attrition rate is approximately 4 percentage points. Comparing the two estimates to the HomVEE attrition standards allowed us to conclude that the sample attrition is within the low to moderate band (acceptable range).

In addition, we checked the extent to which attrition could bias our effect size estimates⁴:

- a- A test for predictors of attrition-- if any outcome and/or family characteristics predict attrition. The findings of the test suggest that the explanatory variables do not well explain/predict attrition. An exception was for the variable that signifies the location of attrition or community is significantly predicting attrition. Households in Bardarash more likely to attrite than households in Gawilan. However, since comparison and intervention groups contain observations from both communities, attrition is proportionally distributed between intervention and comparison. Therefore, even if attrition is associated with community, this does not necessary bias our estimates as attrition is evenly distributed between comparison and intervention groups.

⁴ Additional tests were also conducted, check below.

- a- A test for differential attrition (where a probit model regressed attrition on treatment) suggested no statistically significant difference in attrition between treatment and control groups. In other words, the program participation is not associated/connected with attrition.
- b- A test for difference in outcomes between attritors and nonattritors is statistically insignificant, suggesting that initial outcome levels are not a determinant of attrition.
- c- A test for difference in observed characteristics between attritors and nonattritors is statistically insignificant, suggesting that initial observed characteristics are not a determinant of attrition.

Overall and differential attrition rates are considered low. There was no evidence for a correlation between attrition and household's observed characteristics and outcomes, and between treatment and comparison groups. We, then, use these findings to relax the assumption that missing (attrition) is independent of potential outcomes MIPO. This provides evidence that attrition in our study does not bias our estimates; but it could reduce the power though.

Statistical power

The study design was developed to obtain a power of .80, however, this was affected in our endline data due to attrition and sampling two caregivers belonging to the same household. First, attrition impacts power in a very straightforward way; by reducing the sample size. However, a key driver for power reduction in our design is the fact that, for many households, two caregivers from the same household were sampled (e.g. a wife and a husband). This created a practical difficulty while analyzing the effect of caregiver program participation on children's outcomes, since the effect of the two caregivers from the same household on their children is joint and could not be isolated from one another. Hence, treating those two caregivers as independent observations creates redundancies and would bias the effect estimates on children outcome. On this ground, we used the duplicated data to conduct caregivers; outcome analysis, and to compare the outcomes changes across gender. However, for child-level outcome analysis, we retained only one caregiver in each household where two caregivers are sampled (the caregiver with least non-response "missing data" was retained and the other was dropped). The below table summarizes the changes in power for each target group.

Table 1: Actual Power change compared to the study design

Target group	Hypothesized sample size	Power obtained	Actual sample size	Power obtained	Power loss/gain (percentage points)
A- Caregiver outcomes	250	.80	299	.87	(+) .07
B- Children (3-5) outcomes	250	.80	210	.74	(-) .06
C- Children (0-3) outcomes	250	.80	173	.65	(-) .15

Note: In B and C the numbers are greater than A because of the overlapping observations (caregivers who are associated with one child assessed with IDELA and one with CREDI) and they are counted twice (one time in IDELA analysis and another in CREDI analysis).

The potential and unfavorable implication of reducing the power is decreasing the likelihood of being able to identify an effect size if it actually exists, in other words, maybe we might be unable to reject the null hypothesis when it is supposed to be rejected.

Qualitative Findings

Nearly all respondents in the qualitative study highlighted the difficult life situation that they have to thrive through with their families, sometimes to the extent that some children need urgent medical treatment but their families are unable to provide it due to financial problems. Some of the mostly reported stresses were financial problems, child educational and schooling difficulties and health and medical needs.

In spite of that, almost all of the respondents highlighted the importance of isolating children from these life stresses and support their development through education. Caregivers demonstrated, as well, an understanding of their responsibilities to support their children's development through mentoring, coaching, playing and doing outdoor activities (going to supermarket or seeing relatives). The importance of child education and schooling was an element that was mentioned repeatedly by caregivers. Thus, they prefer to focus most of their child development support on educating their children. Other activities, than education, that caregivers do with their children are such as reading stories, playing sports or exercising and to have open discussions.

Mothers, in the sample, tended to spend more time and to be more engaged with children than fathers; largely because the dads are the main bread-winners and they spend most of their time at work. However, it was also reported that when fathers do not work, they tend to spend more time with their children. As reported by respondents, the program had a positive impact on them since it improved family dynamics, reduced violent child rearing behaviors and induced the ability to better listen to their children and understand their needs. The program was reported to be particularly beneficial for male caregivers as they became more aware of the importance of their involvement in child development and became more engaged in it.

With respect to technique for soothing uneased children, only two out of 14 respondents indicated that they will use a threatening language and maybe beat their children. The reminder of the respondents mentioned that they try to keep their calm and entertain their uneased children through playing, singing or going out for a walk until they are fully soothed.

In terms of unintended impact of the program, respondents reported that the groups created for the training were very helpful and provided psychological and social solidarity. Caregivers were able to connect with other families living under similar circumstances and shared ideas, opinions and discussions about raising children and living situations. In the same manner, the groups also acted as a psychological support to their members through exchanging advises and life testimonies.

Lastly, when caregivers were asked about what type of additional support they need, they indicated financial support in terms of job opportunities and mental and psychological support – particularly reported by female caregivers. Parents also underlined the importance of providing safe areas and playgrounds for their children where they can play and learn at the same time.

Quantitative Results

A- Caregiver and child sociodemographic characteristics

In Table 2 below, we present the average values for a range of socioeconomic characteristics for the parents and caregivers that participated in the survey. More than half of the intervention sample (54% at baseline and 56% at endline) were female respondents, while for the comparison sample, 48% at the baseline and 49% at the endline were female. The average age of respondents ranged between 32 and 33 years old, and on average they have completed 6/7 years of education. Data show that respondents at endline have an average of 21/22 months of residence within the camps. Respondents also displayed similar living situations across intervention and comparison data. There were no other significant sociodemographic differences, suggesting that respondents from comparison and intervention groups have similar background characteristics.

Table 2: Caregiver's sociodemographic characteristics over time

Sociodemographic characteristic	Intervention		Comparison	
	Baseline	Endline	Baseline	Endline
	N=174	N=155	N=205	N=144
% female caregivers	54%	56%	48%	49%
Age	31.9	33.0	33.0	33.1
Education level (last grade completed)	6.9	6.9	6.4	6.9
Months living in the camp	12.5	21.1	13.2	21.8
Household size	5.3	5.7	5.2	5.2
% caregivers that had a partner	95.4%	93.5%	94.1%	95.8%
% of caregivers that lived with their partner	96.9%	97.2%	98.4%	99.2%
Living conditions --SES (out of 9)	4	5	4	5

Around 46-50% of the respondents in our survey had a female child, and the average age of the children in the (3-5 years old) sample was 3.5 years old, while for the younger (0-3 years old) group, the average age was 2/2.7 years old. was 3 years old (Table 4). A low percentage of caregivers and parents reported that their child was enrolled in an ECCD center or similar, but there were changes between baseline and endline, particularly for children part of the older group. Disability for older group was around 14%, and it was between 3-7% for the younger children.

Table 3: Table 3: Child's sociodemographic characteristics reported by the caregiver over time

Sociodemographic characteristic	Children 3-5		Children 0-3	
	Baseline	Endline	Baseline	Endline
	N=263	N=210	N=249	N=173
% female	46%	47%	45%	51%
Age (in years)	3.6	3.9	1.9	2.7
% children enrolled in an ECD center	9%	19%	6.4%	11%
% children with a disability	13.6%	15.2%	3%	6.9%
% children that had at least one storybook	67%	86%	53%	86.7%
Average # of storybooks	2.5	4.5	2.1	3.1

At baseline, around 5 to 7 in 10 caregivers and parents reported that their child had a storybook or an appropriate book for young children. At endline, the majority of caregivers and parents (86-87%) reported that their child had at least one storybook.

Caregivers' outcomes

Caregiver's outcomes are measured by seven indicators (Caregiver Child Relationship, Learning and playing, Caregiver resilience, Caregiver Self-Efficacy, Gender-based attitudes, Gender-based practices and Caregiver perceived Stress). Initially, we began with testing for means difference using a two-sample t test to check if the outcomes vary by gender (Model-1). The analysis indicates a statistically significant difference in the means of four outcome indicators between female and male caregivers who participated in the program. As shown by table (4), the mean differences were higher on two outcomes for female (Caregiver Child Relationship Score and Caregiver Self-efficacy), signifying a higher outcome value for female participants than males. On the other hand, male caregivers had a relatively higher outcome value on Gender-based attitudes and Gender-based practices. These initial findings, however, were considered cautiously since introducing control variables in the DID regression may alter the statistical significance of the means difference.

Table 4: Two-sample t test—Outcomes' mean difference by caregiver's gender (Model-1)

	Mean-M	Mean-F	Diff	p value
Caregiver Child Relationship Score	0.458	.499	-.041	.000
Learning and playing activities	0.651	.702	-.051	.111
Caregiver resilience	0.528	.521	.007	.668
Caregiver Self-Efficacy - 1 child	0.630	.678	-.048	.001
Caregiver perceived Stress Scale	28.870	29.426	-.556	.497
Gender-based attitudes	0.554	.54	.015	.092
Gender-based practices	0.596	.542	.053	.007

To estimate the effect size using the DID method, a linear regression was fitted to the two-periods panel data for each of the seven outcome indicators. The coefficient of the interaction term between period and intervention DID resembles the estimated effect size of program participation on caregiver outcomes. The standard DID's estimating equation of the linear regression (Khandker et al., 2009) is:

$$Y = \beta_0 + \beta_1P + \beta_2T + \beta_3(T \times P) + \epsilon.$$

Where:

P – Period (0 if baseline, and 1 if endline)

T – Treatment (0 if comparison, and 1 if treatment)

$\beta_3(T \times P)$ – The coefficient of the interaction term P*T

In addition, a set of background control variables (location, family size, caregiver sex, age of spouse, live with spouse, length of residence in the camp, educational level and living conditions) were used within the model to improve the effect estimate, by accounting for the variation in outcomes that may have been caused by differences in observed characteristics between the comparison and intervention groups.

The model suggests that the intervention had a positive and statistically significant correlation with four caregiver's outcome indicators (Caregiver Child Relationship, Learning and playing, Caregiver resilience,

Caregiver Self-Efficacy), while for the remaining three outcomes (Gender-based attitudes, Gender-based practices and Caregiver perceived Stress), there is not enough evidence in the data to support a relationship between the intervention and these outcomes. Table (4) summarizes the results of the seven regression models.

Moreover, in the model, the sex of caregiver explains some of the variation on five outcome variables. Female caregivers, for instance, have a higher average score than males on Caregiver Child Relationship, Caregiver Self-Efficacy and Caregiver Perceived Stress Scale. While, male caregivers have a higher average score than females on Gender-based Attitudes and Gender-based Practices. These differences, although being relatively small, they are statistically significant, and confirms the findings from the t-test in Model-1

In the same manner, educational level explains some variation on all the seven outcomes. This is as the higher the educational level of caregivers, the increased likelihood of their scores to be high as well (a positive correlation). These differences are, once again, very small, but they are very statistically significant. Worth noting, family size is, as well, positively correlated with Caregiver Perceived Stress, meaning that caregivers with large family size are likely to report higher perceived stress score.

Lastly, looking at the reported F-statistic, the seven estimated models are statistically significant. However, the models explain a relatively small amount of variation on caregivers' outcome indicators. At best, a model explains roughly 17% of the variation, which would mean that at least 83% of difference in outcomes between caregivers is not observed and cannot be controlled for in the collected data.

Table 5: Diff-in-Diff estimation for seven caregiver's outcome indicator (Model-2)

	(1) Caregiver Child Relationship Score	(2) Learning and playing activities	(3) Caregiver resilience	(4) Caregiver Self-Efficacy - 1 child	(5) Gender-based attitudes	(6) Gender-based practices	(7) Caregiver perceived Stress Scale
0 if baseline and 1 if endline	-0.0654***	-0.0339	-0.0521***	0.00591	-0.00512	0.0304***	0.247
0 if comparison and 1 if intervention	-0.0328***	-0.0483**	-0.0307***	-0.0238**	0.00127	-0.0369***	0.350
DiD	0.0454***	0.133***	0.0513***	0.0316**	-0.00295	0.00498	-0.178
Location	0.0129*	-0.0369*	0.0329***	0.0216***	-0.00993*	0.0165*	-0.611
Family size	-0.000510	0.00188	0.000372	0.000751	0.0000135	0.000153	0.151***
Caregiver sex	0.0361***	-0.0229	-0.0141	0.0719***	-0.0185***	-0.0228***	1.180**
Age of spouse	-0.000201	0.000784	0.000309	-0.000386	0.000794**	-0.000410	0.0161
Live with spouse	-0.0378*	-0.0590	-0.0323	-0.0739***	0.0157	0.0908**	-0.188
Length of residence	0.000146	-0.00169	-0.000768	-0.000297	-0.000469	-0.00207***	0.0127
Highest education	0.00278***	0.0122***	0.00471***	0.00585***	0.00506***	0.00425***	-0.118***
Living conditions	0.0154	-0.00770	0.0532	0.0306	0.0430*	0.0254	-2.317
Adjusted R^2	0.100	0.059	0.073	0.168	0.107	0.085	0.026
F	10.12	5.132	6.807	22.49	9.768	6.873	3.982

* p<0.10, ** p<0.05, *** p<0.01

Further analysis was conducted on caregivers' outcomes to obtain the heterogonous effect size estimate, which is the effect size estimate across different caregivers' characteristics (in our case it is the effect size for male and female caregivers separately). This analysis was undertaken for each of the seven outcomes individually and their findings are presented in the following tables. Male caregivers had a larger and statically significant effect size for three outcomes: Caregiver-Child Relationship, Caregiver Resilience and Caregiver Self-efficacy. Females, on the other hand, had a higher effect on Learning and Playing Activities.

Table 6: Heterogonous effect size estimate per caregiver sex – Caregiver-child Relationship

	(Female) Caregiver Child Relationship	(Male) Caregiver Child Relationship
0 if baseline and 1 if endline	-0.0540***	-0.0748***
0 if comparison and 1 if intervention	-0.0152	-0.0480***
DiD	0.0286*	0.0546***
Location	0.0390***	-0.0147
Family size	0.000978	-0.000524
Age of spouse	-0.00128**	0.000799
Live with spouse	-0.0378	-0.00656
Length of residence	0.0000985	0.000529
Highest education	0.00319***	0.00298**
Living conditions	0.0276	0.0169
Adjusted R^2	0.091	0.089
F	6.148	4.857

* p<0.10, ** p<0.05, *** p<0.01

Table 7: Heterogonous effect size estimate per caregiver sex – Learning and Playing

	(Female) Learning and playing activities	(Male) Learning and playing activities
0 if baseline and 1 if endline	-0.0336	0.0139
0 if comparison and 1 if intervention	-0.0496	-0.0435*
DiD	0.167***	0.0713*
Location	0.00545	-0.0848***
Family size	-0.0116	0.00437***
Age of spouse	0.000671	0.00210
Live with spouse	-0.0566	-0.0940
Length of residence	-0.0000252	-0.00870**
Highest education	0.0120***	0.0116***
Living conditions	-0.00294	0.0327
Adjusted R^2	0.083	0.058
F	4.844	3.887

* p<0.10, ** p<0.05, *** p<0.01

Table 8: Heterogonous effect size estimate per caregiver sex – Caregiver Resilience

	(Female) Caregiver resilience	(Male) Caregiver resilience
0 if baseline and 1 if endline	-0.0511***	-0.0435**
0 if comparison and 1 if intervention	-0.0188	-0.0393***
DiD	0.0393*	0.0507**
Location	0.0562***	0.00808
Family size	-0.00799	0.00147**
Age of spouse	0.000416	0.000990
Live with spouse	-0.0371	0.0380
Lenght of residence	-0.0000318	-0.00229*
Highest education	0.00595***	0.00275*
living conditions	0.0345	0.0832
Adjusted R^2	0.104	0.054
F	5.969	4.892

* p<0.10, ** p<0.05, *** p<0.01

Table 9: Heterogonous effect size estimate per caregiver sex – Caregiver Self-efficacy

	(Female) Caregiver Self-Efficacy - 1 child	(Male) Caregiver Self-Efficacy - 1 child
0 if baseline and 1 if endline	0.00721	0.00183
0 if comparison and 1 if intervention	-0.0104	-0.0362***
DiD	0.00841	0.0544**
Location	0.0366***	0.00233
Family size	0.00349	0.000155
Age of spouse	-0.00119*	0.000263
Live with spouse	-0.0753***	-0.0449
Length of residence	-0.000114	0.0000655
Highest education	0.00567***	0.00697***
Living conditions	-0.00539	0.0754
Adjusted R^2	0.081	0.089
F	6.318	6.027

* p<0.10, ** p<0.05, *** p<0.01

Table 10: Heterogonous effect size estimate per caregiver sex – Gender-based attitudes

	(Male) Gender-based attitudes	(Female) Gender-based attitudes
0 if baseline and 1 if endline	-0.00597	-0.00126
0 if comparison and 1 if intervention	-0.00202	0.00337
DiD	-0.000118	-0.00744
Location	0.000372	-0.0171**
Family size	-0.000154	0.000264
Age of spouse	0.000740	0.000820
Live with spouse	0.0181	0.00270
Length of residence	-0.000517	-0.000648
Highest education	0.00606***	0.00403***
Living conditions	0.0759**	0.0106
Adjusted R^2	0.129	0.049
F	8.071	3.035

* p<0.10, ** p<0.05, *** p<0.01

Table 11: Heterogonous effect size estimate per caregiver sex – Gender-based Practices

	(Female) Gender-based practices	(Male) Gender-based practices
0 if baseline and 1 if endline	0.0386**	0.0218
0 if comparison and 1 if intervention	-0.0365**	-0.0329***
DiD	0.00155	0.00431
Location	0.0160	0.0196*
Family size	-0.00552	0.000544
Age of spouse	-0.00100	0.00115
Live with spouse	0.100**	0.0647**
Length of residence	-0.00208**	-0.00229**
Highest education	0.00451**	0.00305**
Living conditions	-0.0339	0.0941**
Adjusted R^2	0.083	0.061
F	4.496	8.472

* p<0.10, ** p<0.05, *** p<0.01

Table 12: Heterogonous effect size estimate per caregiver sex – Caregiver-perceived Stress

	(Female) Caregiver Perceived Stress Scale	(Male) Caregiver Perceived Stress Scale
0 if baseline and 1 if endline	1.401	-1.362
0 if comparison and 1 if intervention	1.114	-0.462
DiD	-1.477	1.516
Location	-1.356*	0.0953
Family size	0.272	0.133***
Age of spouse	0.0732	-0.0758**
Live with spouse	-0.0149	-2.163
Length of residence	-0.0197	0.101*
Highest education	-0.157**	-0.0632
Living conditions	-2.476	-3.021
Adjusted R^2	0.031	0.010
F	2.251	4.085

* p<0.10, ** p<0.05, *** p<0.01

CREDI: Young Children's (0-36 months old) outcomes

CREDI was used to measure the outcomes of children who are less than 3 years old. CREDI overall score is an aggregate measure of four outcome domains (Cognitive, language, Motor and Social-Emotional). The same regression model used for caregiver's outcome was fitted to 0-36 months children's data. The coefficient of the interaction term "DiD" represents the estimated effect size of program participation on children's outcomes. The regression was fitted two times; onetime for the overall CREDI score, and another time for each of the four CREDI domains.

The regression's outcomes (Table 13) provide no evidence for significant correlation between the intervention and CREDI overall score. Even after controlling for background variables, the estimated effect size remains statistically insignificant at all conventional significance levels (10%, 5% and 1%). This suggests that program participation is likely not associated with any outcomes' changes for children aged between 0-36 months old, who were evaluated by CREDI. A separate regression model was estimated for each of CREDI's four domains. Again, the results, displayed within table (14), do not suggest any significant correlation between the intervention and each of the four domains.

Despite the absence of significant relation between program participation and children's outcomes, CREDI scores reflect changes in the outcomes between the two periods, and these changes were correlated with child age, meaning that CREDI score increases as the child's age increases (statistically significant). This is consistent with the expected results mentioned in CREDI's Data Management and Scoring Manual⁵. This is particularly important as it rules out the possibility of committing a scoring or implementation mistake during the assessment/interviews, and therefore, supports the conclusion that the inability to detect a true effect size is not due to measurement/procedural mistakes.

Table 13: Diff-in-Diff estimation for 0-36 months old children's outcomes (CREDI Overall Score)

	(1) OVERALL	(2) OVERALL
0 if baseline and 1 if endline	0.865**	0.678
0 if control and 1 if treatment	0.0174	0.105
DiD	0.196	0.154
Location		-0.466
Family size		-0.203**
Caregiver sex		0.0784
Age of spouse		0.0226
Live with spouse		-1.482
Length of residence		0.00296
Highest education		-0.0278
Living conditions		2.018*
Number Of Caregivers participated		-0.102
Adjusted R ²	0.028	0.035
F	4.972	2.231

* p<0.10, ** p<0.05, *** p<0.01

⁵ [CREDI Test Administrator Manual \(harvard.edu\)](https://crediharvard.org/wp-content/uploads/2018/07/CREDI-Test-Administrator-Manual.pdf)

Table 14: Diff-in-Diff estimation for 0-36 months old children's outcomes (CREDI Domains' Scores)

	(1) MOT	(2) COG	(3) LANG	(4) SEM
0 if baseline and 1 if endline	0.851*	0.612	0.578	0.766*
0 if control and 1 if treatment	0.103	0.0563	0.161	0.0713
DiD	-0.0626	0.0897	0.200	0.0396
Location	-0.427	-0.252	-0.508*	-0.406
Family size	-0.207**	-0.138*	-0.113	-0.151*
Caregiver sex	0.101	-0.0206	-0.0780	-0.0463
Age of spouse	0.0201	0.0151	0.0206	0.0203
Live with spouse	-1.238	-1.144	-1.387	-1.468
Length of residence	0.00968	0.00692	-0.00246	0.00995
Highest education	-0.0381	-0.0151	-0.0246	-0.0248
Living conditions	1.739	1.302	1.547	1.633
Number of caregivers of the same family participated in the program	-0.0849	-0.00708	-0.0271	0.0109
Adjusted R^2	0.044	0.034	0.034	0.044
F	2.571	2.183	2.203	2.564

* p<0.10, ** p<0.05, *** p<0.0

To understand the connection between CREDI outcomes and caregivers' outcomes, we used a fixed-effect regression model, where CREDI scores (overall and domain scores) were regressed on caregivers' outcomes. This model should allow for identifying any underlying connection between both outcome types. The results suggest that Gender-based attitudes is positively correlated with CREDI's overall and domain scores. This correlation was statistically significant at 5% significance level. Additionally, Caregiver Self-Efficacy is adversely correlated with children's Motor, Cognitive and Social-emotional development, however, the estimates were only statistically significant at 10% significance level.

Table 15: Checking for correlations between caregivers' outcomes and children's outcomes

	(1) OVERALL	(2) MOT	(3) COG	(4) LANG	(5) SEM
0 if baseline and 1 if endline	2.648***	3.055***	2.667***	1.501*	2.624***
0 if control and 1 if treatment	0.294	1.435	1.050	0.474	1.089
Location	0.688	-0.298	-0.0554	0.303	-0.118
Family size	-0.198	-0.647	-0.896	-0.410	-1.097*
Caregiver sex	-0.291	-0.414	-0.429	-0.294	-0.386
Age of spouse	0.159**	0.186**	0.155**	0.141**	0.163**
Length of residence	-0.0980	-0.0955	-0.110	0.00604	-0.0592
Highest education	-0.266***	-0.187**	-0.178**	-0.250***	-0.224***
Living conditions	-1.313	-2.485	-2.658*	-0.851	-2.673*
Number Of Caregivers Participated	-0.212	-0.150	0.0295	-0.0121	0.127
Child 2 age	0.0123	0.0486	0.146	-0.00003	0.134
Caregiver Child Relationship Score	-3.657	-1.816	-1.926	-2.299	-1.568
Learning and playing activities	-0.188	-0.768	-0.707	-0.296	-0.608
Caregiver resilience	1.082	0.914	1.383	0.719	1.155
Caregiver Self-Efficacy	-2.610	-4.143*	-3.463*	-2.888	-4.398*
Caregiver Perceived Stress Scale	0.00808	0.0318	0.0304	0.0111	0.0267
Gender-based attitudes	6.284**	7.657**	7.378***	6.277**	7.095**
Gender-based practices	2.591	1.497	1.909	2.270	2.396
<i>R</i> ²	0.63	0.66	0.61	0.64	0.65
F	5.251	5.980	4.775	5.529	5.891

* p<0.10, ** p<0.05, *** p<0.01

Testing for mean difference of CREDI scores by gender provides no evidence for gender-dependent score difference. In other words, there is no difference between female and male children's average scores.

Table 16: Two-sample t test—Outcomes' mean difference by child's gender

	Mean-M*	Mean-F*	Diff	p value
OVERALL CREDI	50.963	49.764	1.2	.198
MOT	50.635	50.066	.57	.399
COG	50.578	49.754	.825	.103
LANG	51.093	50.208	.885	.139
SEM	50.825	50.148	.678	.252

M* & F* are male and female

IDELA: Young Children's (>3-6 years old) outcomes

For children older than 3 years old until 6 years old, IDELA score was utilized to measure their outcomes. IDELA overall score is an aggregated measure of six indicators (Motor, Emergent Literacy, Emergent Numeracy, Social-Emotional, Executive Function and Approaches to Learning). Similar to the other targeted groups, the change in children outcomes linked to the intervention (effect size) was estimated by fitting a linear regression. The regression was fitted onetime for the overall IDELA score and another time for each of the six IDELA indicators.

The findings of regressing the overall IDELA score, as demonstrated by table (17), suggest a positive and statistically significant correlation between the intervention and children's overall outcomes. In particular, the intervention is correlated with an average increase in the overall IDELA score of 11.05 points, significant at 5% significance level.

To better understand the effect of the intervention on children's IDELA outcomes, the same regression model was fitted for each of the six components of IDELA overall score. The results show a statistically significant positive relation between the program and two outcomes (Emergent Numeracy and Social-Emotional). The program is estimated to have increased both outcomes by 6.5 and 17.5 points respectively. On the other hand, the changes occurred on the remaining four outcomes could not strictly be attributed to program participation since these changes lack statistical significance. Table () displays the results of the six regression models.

Table 17: Diff-in-Diff estimation for 3-6 years old children's outcomes (IDELA Overall Score)

	(1) Total IDELA	(2) Total IDELA
0 if baseline and 1 if endline	9.044***	10.73**
0 if comparison and 1 if intervention	0.0314	0.212
DiD	10.16**	11.21**
Location		3.194
Family size		-0.174
Caregiver sex		5.057**
Age of spouse		-0.151
Live with spouse		3.505
Length of residence		-0.0798
Highest education		0.303
Living conditions		-1.913
Child sex		0.153
Attend Kindergarten		1.582
Child age		-2.178**
Number of caregivers participated		3.165
Adjusted R^2	0.107	0.115
F	19.95	4.827

* p<0.10, ** p<0.05, *** p<0.01

Table 18: Diff-in-Diff estimation for 3-6 years old children's outcomes (IDELA Domains' Scores)

	(1)	(2)	(3)	(4)	(5)	(6)
	Motor	Emergent Literacy	Emergent Numeracy	Social- Emotional	Executive Function	Approaches to Learning
0 if baseline and 1 if endline	17.01	17.82**	2.540	5.566	10.57	10.71***
0 if comparison and 1 if intervention	0.731	0.398	-0.302	0.0195	-0.212	-0.0585
DiD	8.713	11.84	6.810***	17.47***	6.727	-0.785
Location	6.061	9.397**	1.017	-3.700	-0.800	3.834*
Family size	-0.229	-0.176	-0.0513	-0.241	-0.222	-0.0986
Caregiver sex	5.503	9.429**	1.442	3.856	9.832**	3.607**
Age of spouse	-0.200	-0.258	-0.0973	-0.0488	-0.339	-0.138
Live with spouse	7.008	5.012	-2.584	4.583	5.284	0.833
Length of residence	0.0566	-0.529	0.0165	0.136	-0.152	-0.184
Highest education	0.260	0.531	0.245*	0.174	0.832*	0.168
Living conditions	14.09	-11.11	-7.884	-2.748	4.183	2.322
Child sex	2.355	-0.501	-1.542	0.299	-0.561	-1.662
Attend ECD	-1.430	8.186	0.0522	-0.481	-0.426	-0.591
Child age	-2.662	-4.481***	-0.161	-1.406	-4.862***	-1.502**
Number of caregivers participated	2.949	7.314*	1.351	1.047	-0.240	4.771***
Adjusted R^2	0.028	0.074	0.060	0.079	0.039	0.056
F	1.858	3.360	2.881	3.514	2.189	2.763

* p<0.10, ** p<0.05, *** p<0.01

Additional analysis was conducted on children's outcomes to obtain the heterogonous effect size estimate. In this case, we checked for heterogonous effect by child's sex and caregiver's sex. This analysis was undertaken for IDELA overall score. According to the results, female children had higher effect size than male children, significant at 10% significance level. Male children, however showing an effect size, their estimate was not statistically significant. In similar ways, comparing the effect sizes between female and male caregivers shows that children of female participants displayed on average a higher and statistically significant (5%) effect size than children of male participants, whose results were not significant.

Table 19: Heterogonous effect size estimate per child's sex – IDELA Overall Score

	(Female) Total IDELA	(Male) Total IDELA
0 if baseline and 1 if endline	-4.732	13.34***
0 if control and 1 if treatment	0.844	0.903
DiD	14.30*	6.859
Location	4.635	1.810
Family size	1.271	-0.197
Caregiver sex	6.896	3.739
Age of spouse	-0.113	-0.259
Live with spouse	9.813	0.698
Length of residence	1.540	-0.193
Highest education	0.809	-0.0371
Living conditions	13.05	-10.18
Child age	-3.200*	-1.648*
Attend ECD	-1.769	4.757
Number of caregiver participated	2.440	4.992*
Adjusted R^2	0.097	0.130
F	2.578	3.515

* p<0.10, ** p<0.05, *** p<0.01

Table 20: Heterogonous effect size estimate per caregiver's sex – IDELA Overall Score

	(Female) Total IDELA	(Male) Total IDELA
0 if baseline and 1 if endline	11.33	10.70**
0 if control and 1 if treatment	1.191	-1.801
DiD	13.40**	8.131
Location	5.549	0.441
Family size	-0.218	0.00145
age of spouse	-0.122	-0.203
live with spouse	4.717	-0.152
Length of residence	0.423	-0.513
Highest education	0.270	0.419
Living conditions	-3.210	-2.435
Child age	-5.426**	3.240**
Attend ECD	6.426	-5.625
Number of caregivers participated	7.180*	1.396
Adjusted R^2	0.128	0.114
F	4.050	2.706

* p<0.10, ** p<0.05, *** p<0.01

Similar to children assessed by CREDI, we implemented a fixed-effect model to check for a direct connection between caregivers' outcomes and IDELA's overall and domains' scores. The only significant correlation we could identify is between Caregiver-children relationship score and children's Emergent Literacy. The estimate is significant at 10% significance level.

Table 21: Checking for correlation between caregivers' outcomes and children's outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Total IDELA	Motor	Emergent Literacy	Emergent Numeracy	Social- Emotional	Executive Function	Approaches to Learning
0 if baseline and 1 if endline	18.44**	22.33	25.96*	9.719*	15.77	17.27	8.043
0 if control and 1 if treatment	-7.719	-15.20	-8.042	-1.255	-6.379	-5.586	-5.426
Location	-5.848	-0.232	-11.27	-3.287	-8.600	7.850	0.0335
Family size	-0.364	-0.422	-0.713	-0.267	-	-0.0987	-0.124
Caregiver sex	8.602	11.21	19.35*	3.315	0.0540	10.93	12.02**
age of spouse	-1.311	-1.722	-2.180	-0.946*	-0.396	1.194	-0.779
live with spouse	-3.151	2.871	-9.349	-2.902	-3.222	1.957	1.532
length of residence	0.180	1.188	-0.236	-0.00217	-0.228	0.135	0.292
Highest education	0.0365	-0.656	0.178	0.771	-0.147	1.454	0.466
living conditions	-19.81	-42.77	-24.75	-18.47	6.729	-24.41	-6.187
child age	-1.688	-4.293	-3.494	-3.045	4.080	-7.952	1.068
Number Of Caregivers Participated	5.029	-4.008	19.74**	0.687	3.697	7.387	8.016*
Caregiver Child Relationship Score	-39.64	-40.47	-94.12*	-27.41	3.431	10.06	-3.185
Learning and playing activities	1.328	4.443	-19.96	4.966	15.86	-6.248	-7.058
Caregiver resilience	23.93	40.79	34.34	3.335	17.25	-44.55	13.35
Caregiver Self-Efficacy - 1 child	4.950	13.18	-14.77	6.196	15.19	2.519	-12.70
Caregiver Perceived Stress Scale	-0.0833	0.0731	-0.360	0.0613	-0.108	-0.900	0.115
Gender-based attitudes	31.25	78.19	21.50	8.887	16.42	66.85	41.02
Gender-based practices	-22.76	-64.26	22.45	-5.287	-43.94	19.05	-7.626
R^2	0.19	0.11	0.15	0.16	0.15	0.095	0.16

F	1.931	1.012	1.433	1.585	1.401	0.849	1.556
---	-------	-------	-------	-------	-------	-------	-------

* p<0.10, ** p<0.05, *** p<0.01

Testing for mean difference of IDELA scores by gender provides no evidence for gender-dependent score difference. In other words, there is no difference between female and male children average scores.

Table 22: Table 16: Two-sample t test—Outcomes' mean difference by child's gender - IDELA

	Mean-M	Mean-F	Diff	p value
Total IDELA	17.008	22.8	-5.792	.454
Motor	23.249	30.345	-7.096	.66
Emergent Literacy	14.853	27.951	-13.098	.313
Emergent Numeracy	9.136	7.563	1.573	.729
Social-Emotional	20.791	25.338	-4.547	.641
Executive Function	10.326	13.758	-3.433	.757
Approaches to Learning	7.942	8.149	-.207	.966