Third Impact Assessment Survey Report on the CONAFE "Early Education" Program

Table of Contents

List of Tables 4
Presentation of Report 6
Part I. Data Gathering Processes and Descriptive Analysis 9
1. Instruments 9
2. Training 12
2.1 ASQ-3 Application Training 14
2.2 Basic Implementation Protocol and Context Instruments Training 16
2.3 Fieldwork and Personnel Selection 18
3. Survey 20
3.1 Supervision 22
3.2 Indigenous Languages 23
4. Capture and Validation 24
5. Descriptive Statistics in Panel Surveys 25
5.1 Home Response Rate 25
5.2 Attrition on Caregivers and Children 31
5.3 Home 32
5.4 Caregivers 41
5.5 Pregnant Women 50
5.6 Fathers 54
Part II. Impact Analysis of Primary Caregivers and Children and Balancing of Groups 58
6. Take Up 58
7. Impact on Caregiver Practices 59
8. Impact on Child Development 62
9. Robustness and Heterogeneous Treatment Effects 63
10. Balance between the Treatment and Control Groups 65
10.1 Household Balance 65
10.2 Caregiver Balance 68
10.3 Pregnant Women Balance 71
10.4 Balance of Fathers 73
10.5 ASQ Balance 74
Conclusions 76

Appendix I. Third Wave of the Panel Study Work Plan to Evaluate the Impact of the Early Childhood Education Program 77

Part I 77

Appendix I. Third Wave of the Panel Study Work Plan to Evaluate the Impact of the Early Childhood Education Program. 80

Part II 80

List of Tables

Table 1: Training Activities 14
Table 2: Survey Staff Information 19
Table 3: Indigenous Languages and Dialects by State 24
Table 4: Communities by Federal Entity 26
Table 5: Relative Response Rate in Households by Federal Entity and Survey Waves 27
Table 6: Interview Results 28
Table 7: Relative Response Rate in Households from Second and Third Survey According to Interview Results and Type of Community 29
Table 8: Interview Results by Federal Entity 30
Table 9: Attrition Descriptive Analysis 31
Table 10: Attrition Regression Analysis 32
Table 11: Is the person being interviewed the same person in this household that was interviewed during the previous survey done by CIDE? 33
Table 12: Children under 60 months of age in the household 33
Table 13: LANGUAGES Does [….] speak Spanish? Members 5 years of age or older. 34
Table 14: Does [….] know how to read and write a message? Members 10 years of age or older. 34
Table 15: Highest level of education completed regarding members 17 years of age or older 35
Table 16: Of what material is the majority of the floor in this house made? 35
Table 17: Of what material is the majority of the walls of this house made? 36
Table 18: Of what material is the majority of the roof of this house made? 37
Table 19: During the last week, how many days was power available for at least three hours per day? 37
Table 20: To which type of drainage or sewer systems is this house connected? 38
Table 21: Appliances, goods and services in the homes. 40
Table 22: What relationship do you have with […..]? 43
Table 23: Is someone in […..]'s household currently a recipient of the Oportunidades Program? 43
Table 24: In […..]'s household, has anyone received support from: Seguro Popular, Procampo, Scholarships or educational programs for senior citizens? 44
Table 25: In […..]'s household, has anyone received support (such as cash, remittences, clothing, foodstuffs, etc.) from people that live in another country or in Mexico during the last year? 44
Table 26: Approximately how many books does […..] have? 45
Table 27: How often you have the opportunity to read or tell stories to […..]? 46
Table 28: Some parents dedicate time to teaching their children, while other parents believe that children learn better on their own. Which of the statements below best describe what you think? 47
Table 29: Sometimes children are well behaved and sometimes not. During the last week, approximately how many times, if it occurred at all, did you have to spank […..]? 48
Table 30: I am now going to ask you to please tell me if the following activities or situations occur: 49
Table 31: Up to which level of education do you believe […..] will study? 49
Table 32: Number of pregnant women in the home 50

Table 33: During pregnancy, did the father of the child accompany you to a medical exam? 51
Table 34: During pregnancy, did you take any medicine or remedy that was not prescribed by a physician? 51
Table 35: Over the last 7 days, how often did you …? 52
Table 36: Is the food you are currently eating different from what you were eating before getting pregnant? 52
Table 37: Compared to what you were eating prior to the pregnancy, you are now eating …. 53
Table 38: Do you believe infants are aware of what is going on around them? 53
Table 39: Up to which level of education do you believe your child will study? 54
Table 40: How often does [….] eat together with you and his mother? 55
Table 41: I am now going to ask you to please tell me if the following activities or situations occur: 56
Table 42: Health risk behaviors, socially and within the home 56
Table 43: Up to which level of education do you believe [….] will study? 57
Table 44: Intermediate results: take-up 58
Table 45: Caregiver Attendance at meetings, first year and second year 59
Table 46: Impact on Observed Parenting Practices – 8 practices + an index – age (0-3) 60
Table 47: Impact on Observed Parenting Practices – 12 practices + an index – age (3-6) 60
Table 48: Impact on Reported Parenting Practices – Brief Summary (sign/significance)- Age (0-3) 61
Table 49: Impact on Reported Parenting Practices – Brief Summary (sign/significance)- Age (3-6) 61
Table 50: Impact on Child Development 62
Table 51: Heterogeneous Effects on Child Development Based on Gender of Child 63
Table 52: Heterogeneous Effects on Child Development Based on Baseline Age of Child 64
Table 53: Differences Between Control and Treatment Households According to Panel Survey's Waves 67
Table 54: Differences Between Caregivers in Control and Treatment Households According to Panel Survey's Waves 70
Table 55: Differences Between Pregnant Women in Control and Treatment Households According to Panel Survey's Waves 72
Table 56: Differences Between Fathers in Control and Treatment Households According to Panel Survey's Waves 73
Table 57: ASQ results 30 months per group for the third survey 74
Table 58: ASQ results 42 months per group for the third survey 75
Table 59: ASQ results 54 months per group for the third survey 75

Presentation of Report

The Center for Research and Teaching in Economics and the World Bank has carried out an impact assessment since 2011, with funding from the International Initiative, for Impact Evaluation (3ie), whose primary aim was to estimate the effects of "Early Education" program intervention implemented by the National Council for Educational Development (CONAFE). This study sought to generate rigorous and useful evidence for decision-makers, researchers and policy makers mainly in the education sector.

The study was conducted through the implementation of a comprehensive three-stage longitudinal data collection survey: baseline, second and third stages of the study. To report the status of research, summarized reports were generated upon completion of each of the first two stages of information gathering. The data collected at the baseline stage were presented with the primary aim of reporting on the design and assessment activities at the first stage of the project. Given the experimental methodological design of the study, one of the important findings at this stage was that the treatment and control groups were statistically similar in contextual variables, child-rearing practices of the parents and skill development of children under three and a half years of age.

The report generated for the second wave of the panel had a similar overall objective: describe the information gathering processes. Among other important variants provided throughout this document, we now have an updated view of the two surveys. Particular emphasis was put on the comparison of the results of both surveys, as well as in estimating the effects of preliminary impact on the main variables of interest in the research and early childhood education program: the cognitive development of children and the child-rearing practices of the principal caregivers. Also, two qualitative data collection surveys were conducted to complement the available data related to the implementation of the intervention in the communities.

Three important conclusions were drawn out of the second stage. The first was that, given the non-response rate, the treatment and control groups remained balanced in the important contextual variables; this situation both supported and validated the experimental method used for the evaluation. The second was that significant changes were seen in caregivers practices directly observed by enumerators and one of the five areas of children development. Note that even when impact results were preliminary, the tendency, even in other variables, were positive. The third was derived from the qualitative surveys which enabled testing of intervention implementation hypotheses and helped explain different results found in the quantitative estimates.

In this third and final report, emphasis will be placed on two general aspects important to this assessment. On the one hand, we present a comparison of study results at each of the stages by describing the following elements: the processes of the third data collection, and the main departures from the first two stages; and present descriptive statistics for this last phase of the study and compare them with those of the previous two stages. On the other hand, we present the impact estimates for child-rearing practices of the main caregivers and the cognitive development results of children at all three stages.

Given the objectives, the document is divided into two major sections. On the one hand, there is the first section in which data gathering processes and descriptive analysis are mainly attended. On the other hand, the second major section focus is placed on presenting impact estimates derived from this assessment and the balance between treatment and control groups. The processes are divided into five subsections. In the first subsection, data collection instruments and their main changes for this third stage are described. In the second, we report on the training process and the subsequent adjustments implemented to improve results. In the third, we present the implementation and operation of the data collection process and the secondary processes. The capture and validation processes are described briefly in the fourth subsection. In the fifth, major descriptive statistics and analysis of the effects of non-response rate on the estimated impact of the evaluation are presented.

The impact section, meanwhile, continues with five more subsections. The sixth deals with matters relating to the operation of the intervention and its importance in giving clarity to the estimates and impact results. The seventh and eighth subsections are intended to report on the impact identified in the child-rearing practices of primary caregivers and the

cognitive development of infants, respectively. In the penultimate subsection, the analysis of heterogeneous effects of the intervention treatment is presented with their degree of robustness by observing the data by sex and age of the infants. The last subsection focuses on reporting on the balance kept in the treatment and control groups. Finally, some conclusions from the analysis made throughout this document are discussed.

Part I. Data Gathering Processes and Descriptive Analysis

1. Instruments

In this third stage of data collection, as in the previous, it was necessary to design an adjusted version of the data collection instruments. The main objective was to respond, according to their impact on quality of information, to two types of differences that arose during this longitudinal study: the substantial differences that arise naturally from the temporality of the surveys; and those related to protocol alignment developed in order to make a corresponding selection of panelists. Additionally, however, due to meetings with CONAFE officials and researchers from the World Bank during the design stage of the third survey, another target to be observed was suggested for integration into the questionnaires, in addition to ASQ development tool measures: infant temperament measures that are contained in the Child Behavior Questionnaire (CBQ).[1]

Moreover, it is important to recall the instruments and their variants that were designed for the study. Initially five different questionnaires were used, although in the last of the three surveys six were employed. The diversity of instruments is based on the variables of interest in the study and the type of individual recipients attending early education intervention sessions; namely primary caregivers, who can be mothers, fathers or other figures; pregnant women; fathers; and infants. Only four of the six instruments were designed by researchers from the CIDE and World Bank team: Household, Primary Caregivers[2], Pregnant Women and Fathers. The Household questionnaire, the protocol with which the interview is opened, collects data about household members and housing features. The Caregiver survey mainly explores the home environment in which children develop, while the pregnant women questionnaire focuses on the habits of the mother during pregnancy and early infant care. Through the Fathers questionnaire, data is

[1] Both instruments, ASQ and CBQ, were developed independently by different researchers and institutions. The ASQ questionnaires have the following references: Jane Squires. Ages & Stages questionnaires. 3rd ed. Baltimore, Md.: Paul H. Brookes, 2009. The children's behavior questionnaire had the following references: Rothbart, M.K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at 3-7 years: The Children's Behavior Questionnaire, *Child Development*, 72, 1394-1408.

[2] The Primary Caregivers Questionnaire is a hybrid from a design perspective given that CIDE researchers designed its structure, but this was not the case in all of the items of which it is comprised. The most important sections for gathering data on primary caregiver child rearing practices was taken from two sections of the HOME Short Form Questionnaire: children from 0 to 35 months and children from 36 to 72 months. The questionnaires can be reviewed online at the National Longitudinal Surveys (NLSY79) website: https://wwww.nylsinfo-org

collected on the social behavior and degree of involvement in the child's development on the part of the male parent.

The fifth is the questionnaire, "Ages & Stages", is an instrument that was not designed by the team, but is extremely important because it is used to measure one of the main constructs of the evaluation: cognitive development of infants. ASQ instruments use twenty versions whose nature responds to the different age ranges (in months) of the infants[3]. The sixth instrument is the Children's Behavior Questionnaire (CBQ) and, as already mentioned, was aimed at measuring temperament levels and some social-emotional characteristics of infants through caregiver responses. It is worth mentioning that, except for the twenty versions of ASQ, all questionnaires used (Home, Caregiver, Pregnant Women, Father and CBQ) were designed with two variants: one for collecting information on treatment communities and the other for control communities.

Delving into the modifications for this stage, we first point to the analysis and implementation of related background data changes corresponding to the substantial ones, as previously mentioned. The objective was to make changes that would otherwise impact information collection items where reference periods are required. The significance of these changes was in the need to obtain precise measures of discrete data of interest in all questions where intervention cycles are specified and survey times are required.

A representative example is seen in questions where information on the number of sessions were attended by intervention recipients and why it was necessary to determine the period that defines attendance, both in treatment and control communities. Items of this type are found in the Caregivers and Household questionnaires. Periods, in these cases, had to be adjusted so that it was clear to the interviewer that the respondent did not provide overlapping information or data combined with information from the previous two surveys.

The other kind of changes addressed the need for accuracy in the identification of respondents within the household, such as major changes to interview protocol. In these cases, all items where information was required from a household member who knows the data of interest were changed including those in the primary caregivers, pregnant women and fathers' questionnaires. Here adjustments are mainly seen in items that serve as a

[3] The age ranges are: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 27, 30, 33, 36, 42, 48, 54 and 60 months.

filter to ensure that the respondent, in any of these questionnaires, provided the appropriate information.

This last type of modification made is important in two ways. The first is that it retains the accuracy of the information, since it contributes to define and select the people who have actually responded in previous surveys of this project more accurately. The second revolves around avoiding unnecessary costs of questionnaire field researchers by working only with individuals who are part of the study itself from baseline, and leaving aside interviews unnecessary for assessing individuals. In short, modifications and redesign of instruments were important parts of maintaining the quality of the data collected at all stages.

2. Training

Training for the third survey was carried out over a period of five days from July 7th to 11th, 2014. A particular advantage in this process was that recommendations from field diagnostics in both previous stages were integrated into the main objective of the training. Three particular suggestions were taken into account in designing and operating the training process: improve the treatment provided by the enumerator when interviewing people and talking to CONAFE staff during field studies, select the optimal amount of interviewers with ability to implement testing with children, and consider variants of context that could arise in the field, especially those regarding languages other than Spanish spoken in some communities.

Based on this analysis, five key measures were taken. The first was an agreement with the company to pre-select the field interviewers to participate in the training process. The second was to allocate more and better distribution time for the ASQ training process, thereby placing a heavier workload of these instruments on interviewers. The third focused on redesigning the survey taker manual, which was simultaneously presented and reviewed during interviewer protocol instruction with an explanation of all questionnaires. The fourth step involved the integration of six CIDE supervisors throughout the training whose main functions were to support, evaluate and select the interviewers with the best profile for data collection. The last step focused on the selection of locations.

The company was responsible for recruiting staff experienced in similar projects; however, this activity was done in conjunction with the CIDE team. Thus, interviewers familiar with panel survey procedures for direct interviews in households were selected. Through a verification process that uses historical database information from previous survey projects, the list of survey takers proposed by the company was analyzed by CIDE research team staff. Based on this process it was possible to rule out survey takers with attitude problems identified in the field or with insufficient ability to follow the complex survey protocols of this project.

Additionally, the company agreed that there would be no further training or replacement of staff. Thus, only personnel initially proposed by the company and validated by the CIDE could participate in the final selection process of survey takers, in other words, no survey taker interested in participating was allowed to miss a single day of training, nor was it possible to add staff.

It should be noted that training, as in the previous two surveys, was conducted on CIDE premises where a total of 35 field researchers from the company were attended, including supervisors.

In general, the training process is structured as follows. The first day, the purpose, scope and importance of the research project was taught. As well, an in depth explanation of the survey content constructs was also provided, particularly those related to the relevant variables of the project: cognitive development of children and parenting practices. Finally, the process focused on the ASQ tools and application practices, activities where the most time is spent. The session closed with a period to answer specific questions (Table 1).

On the second day, the entire survey protocol and interviewer manual was reviewed. An activity of vital importance on this day was to review the context questionnaires, especially the part of the protocol on selecting the correct respondent by type of questionnaire. Once a session of questions about the interviewer manual was closed, the application protocol and agreements generated during training both days, arrangements of the survey logistics continued. The logistics agreed upon in conjunction with the company managers was reviewed by CIDE supervisors in order to establish the coordination guidelines required to fully comply with the preset field schedule.

Table 1: Training Activities

Days of the Month July, 2014	Contents
Monday 7 *CIDE Classroom*	1. General purpose of the project. 2. Child development concepts. 3. ASQ assessment instruments.
Tuesday 8 *CIDE Classroom*	1. Presentation protocol and data collection method. 2. Basic data collection tools. 3. Field conduct protocol.
Wednesday 9 *Activity*	Practice in the State of Hidalgo. Data collection with supervision regarding protocol and tool control.
Thursday 10 *Activity*	Classroom feedback regarding field observations, response to questions and final selection of survey staff.
Friday 11 *Classroom*	Practice in the State of Mexico. Data collection with supervision regarding protocol and tool control.

2.1 ASQ-3 Application Training

The program on child development and familiarization with the Ages & Stages Questionnaire (ASQ) testing was conducted by a specialist trained in child psychology at the National Institute of Perinatology and experienced in the application of ASQ tools with groups of Mexican family parents. Previously it was noted that during the first day, an explanation of the most important constructs of the evaluation was given; additionally, age ranges and their relevance to children tests were taught.

Following this activity, the types of questions found in ASQ questionnaires were covered in depth. Once several conceptual questions were answered, the teaching materials[4] used during the evaluation were distributed and practices were conducted. As in previous

[4] Teams of two interviewers were formed, and each team was given a closed bag with the teaching material required for ASQ implementation. The material is as follows: two small wooden blocks and eight large wooden blocks, a puzzle (the same model for all), a cloth with button closure, a booklet with images laminated in plastic, cereal loops, a small plastic bottle, a 500ml plastic bottle, a pair of scissors (children's), a cup (3.5 inch Peruvian) a rag doll, a small ball (size2-sponge), a medium ball (size 10, plastic), a mirror (plastic), six crayons, a shoelace, five bottle caps with hole, a toy car (plastic), a spoon (plastic), a rattle, a Ziploc bag (17x20), a Ziploc bag (27x28), a bag of toys (raffia), a watercolor marker, a story book, 40 white sheets, a pack of cleaning towels (to be used to clean equipment after each interview). All materials were replaced by new items for the third survey given their degree of wear.

training, an explanation of the three question codes found in ASQ questionnaires -A, B and C- was given, where emphasis was put on the implementation of type C. This type of multiple choice questions are the most delicate given that they require interaction between the child, the primary caregiver and field researcher, the very reason why special attention during the training is required. The three codes are described below:

- **Code A**. Assigned to questions that are directed to the primary caregiver that do not require the performance of any activity.
- **Code B**. Assigned to questions requiring an action by the child and the respective observation by the fieldworker. Questions with this code state that if the child does not perform the action in front of the researcher, the response is expected to be provided the primary caregiver.
- **Code C**. Questions assigned this code indicates that the child must perform an action and the researcher, meanwhile, makes an assessment solely on the basis of the observed response. Occasionally, however, due to force majeure, situations are faced where it is only possible to ask the primary caregiver. Among these situations are absences of the child due to illness, extreme anxiety noted in the child's behavior during the interview or simply the infant does not want to participate.

A determining factor for the quality of instruction information is proper protocol implementation in calculating the age of infants. They are twenty different versions of ASQ, each of which corresponds to one of the different age ranges of the children. Therefore, the calculation method is extremely important because an accurate assessment of cognitive development of a child necessarily depends on the person who applies the questionnaire according to the age range of the child. A miscalculation in this process could create a large bias in two ways: the overestimation of the cognitive development of a child, on the one hand, or understatement on the other. Therefore special attention in this process-related training was provided.

Finally, the interviewers were instructed on the code of ethics which is aligned with the tools implementation protocol, especially regarding the interaction required with both the child and the primary caregiver. Also explained, as in previous surveys, was that the protocol is based on respect for the primary caregiver, as well as all household members, regardless of their beliefs, origins, customs and habits in relation to the child.

2.2 Basic Implementation Protocol and Context Instruments Training

Conversely, training on general application protocols, was given by research team members and was primarily focused on four aspects: review of all context instruments, the survey application process, connection with the operation of the evaluated intervention and a further review of the most important issues of the CIDE ethics code. For good communication, the way in which the instruments[5], not just the context, are aligned with the type population to which the intervention is aimed:

- Instrument for Primary Caregivers: aimed at those people who spend the most time with the infants.
- Instrument for Pregnant Women: aimed at women in this condition in order to learn about prenatal care and baby care.
- Instrument for Fathers: aimed at fathers to inquire about their parenting practices.
- ASQ and CBQ instruments: aimed at gathering cognitive development data and information on the behavior of children from two points of view. The ASQ considers two primary sources: the perspective of the primary caregiver (items A and B) and by direct observation of activities (item C). The CBQ, meanwhile, is aimed at caregivers only, without direct activities with the children.

Three important concerns related to the CBQ questionnaires should be mentioned. First was to emphasize that the entire questionnaire is administered directly to caregivers and activities with infants are not required. As such, during the training the questionnaire was classified as just another context questionnaire. Consequently, and contrary to the ASQ training, a specialist was not required and the CIDE staff team was able to give the training. Finally, they used the shortest version of this questionnaire, which was translated into Spanish.

Moreover, a brief explanation of the CONAFE intervention was given in terms of activities and major players. One of the focus points was to instruct survey takers on the type of

[5] Although the Home questionnaire is considered a context instrument, it is not directed to a particular respondent figure as this person may be any member of the household not less than 15 years of age and who knows the information.

sessions that are held in treatment communities and how they, in turn, serve depending on the classification of the people receiving the intervention (treatment) and in control communities. Similarly, the CONAFE representative and her position in the operational chain of the program was introduced, and the importance of that role for the inhabitants of the treatment communities. The importance of data consisted in the need of the survey takers to interact with the CONAFE representatives during the information gathering.

Likewise, the CONAFE area supervisors and their degree of responsibility in relation to the operation of the program were mentioned, focusing on their hierarchical position and number of representatives (up to ten representatives). During the training, the general concepts of *impact assessment and random experiment* were covered, as well as the purpose of the assessment and that in this case it was to document the effects of CONAFE early education program.

Mentioned was that the assessment of early education program, at this stage, would require the use of at least one and up to four general data collection and cognitive development tools (ASQ and CBQ) to gather the information, thus it was necessary to explain their content and sequence of application in both administration formats: treatment and control groups. This way the relationship between the intervention, the number of members of the same household who attend sessions or attend CONAFE preschool education and the amount of instruments to manage per household was made clear.

Furthermore, an introduction to the kind of items, their maesures and the way in which they are related to the impact assessment was conducted. Issues such as the complementary role of open questions in an instrument, and the type of closed items with dichotomous, polychotomous and discrete data reception options were discussed. The importance of reviewing the survey periods was emphasized for open questions related to discrete data.

Prior to the protocol training closure, CIDE policies concerning the field research code of ethics were discussed. Special care was placed on the diversity of customs found in the villages throughout the different entities to be visited. With help from a child psychology specialist, the importance of dealing with children during the application of ASQ questionnaires was reiterated. Finally the analyses of previous surveys were presented with the aim of emphasizing punctuality and the clothing required to work properly in the field.

2.3 Fieldwork and Personnel Selection

In order to apply what was taught in the research work training, two days of practice were held in different locations. This time, federal CONAFE authorities coordinated the visits company and CIDE researchers made to households currently participating in the Early Childhood Education program but not part of the sample (treatment) and households that do not receive intervention (control).

The town of Calnali in the State of Hidalgo was visited on July 9th. The visit to the Calnali site made a great difference in the training as the population speaks a different language. This way it was possible to interview households in a single location with the following characteristics: 1) households with intervention - treatment; 2) households without intervention - control; 3) both types of households with bilingual –Spanish and an indigenous language spoken; and 4) households where only an indigenous language was spoken. Due to this visit, field researchers were able to pose interesting questions which were subsequently answered, thus improving the use of the tools.

On the fourth day, a dynamic feedback session was held at the CIDE facilities with the aim of gathering all of the questions that had arose from the protocol implementation in Calnali. In previous surveys, field practices were conducted over two consecutive days, however, this time the questions from the field were resolved first. This change in the flow was conducted so that the final day of practice, the researchers would be able to hold the interviews after having cleared their doubts, which in turn, CIDE supervisors had more accurate and expedient information to assess survey takers on the last day of training. On July 11th, the town of Ocoyoacac in the State of Mexico was visited to conclude the last field practice.

Table 2: Survey Staff Information

	Survey Taker			Company Supervisors			CIDE Supervisors			Total/State		
Wave	I	II	III	I	II	III	I	II	III	I	II	III
State												
Chiapas	12	10	4	1	1	1	1	2	1	**14**	**13**	**6**
State of Mexico	7	16	8	2	1	2	3	2	1	**12**	**19**	**11**
Oaxaca	4	4	4	1	1	1	1	2	1	**6**	**7**	**6**
Puebla	8	12	4	2	2	1	1	1	1	**10**	**15**	**6**
Veracruz	16	16	3	1	1	1	2	1	1	**19**	**18**	**5**
Queretaro	13	12	4	0	3	1	2	4	2	**15**	**19**	**15**
Total number of survey takers and supervisors per survey – company only –										**44**	**65**	**35**

Note: The total (Total/State) number of survey takers and supervisors per wave – company only – do not correspond to the sum of all rows because the surveys of the various entities were not performed simultaneously. Given this situation, the surveyors who visited a given state could also visit others and thus would be counted more than once in each of the rows.

Note: The Arabic numerals "I", "II" and "III" refer to the first, second and third surveys

At the end of the last day of training, the CIDE supervisors were asked to select the field interviewers who would actually participate in the survey. Among the evaluation criteria for the final selection, the following were considered: knowledge level of data collection instruments, interview protocol and code of ethics, ability to deal with children, team work skills and punctuality. After analyzing their profiles together with the field monitoring results, a total of 28 interviewers were selected. The company supervisors, meanwhile, were selected internally without the participation of the research team.

3. Survey

The first agreements required to carry out the survey, including logistics, were made during the first week of July. The members of the research team and polling company staff (Berumen y Asociados) went to the CONAFE state delegations to agree on work paths. The design of the final work plan was completed on July 10th and included adjustments to the dates agreed in meetings with CONAFE representatives. The schedule included the following information for each of the seven work paths: state, county and town; number of households to visit in the community; number of interpreters, if necessary; contact information for control communities or the CONAFE supervisor in treatment communities; comments regarding the survey, path number and visitation dates.

The CIDE/BM[6] research team simultaneously provided the company with a database containing identification information and geographic location of each of the households in the sample. Delivered variables were: state, county and town, type of location (treatment and control), folio number of households, identification of individual household members, address (street, number and neighborhood), description of the property, names of household members recorded in both previous surveys and identification of persons to be interviewed by type of survey questionnaire: anchor child, caregiver, pregnant woman and father, (the last two if applicable) [7].

An important aspect in relation to the care of the data was that the CIDE/BM team again gave the company all directory information on pre-printed labels. Also, instructions were given that the household questionnaire labels should be adhered in the field prior to the interview. This strategy avoided errors in the recording of information about the home; it should be noted that it was expected that these data were the same considering that the study had a relatively short duration of two years. Either way, in cases where the data had changed (change of address, for example), it was requested that the new information be recorded in the space provided in the questionnaire and the setting be indicated.

---

[6] The research team is comprised of both researchers and assistants from the Center for Economic Research and Education (CIDE) and the World Bank (WB).

[7] The anchor child is the one who at baseline was assigned as the main reference for managing the household questionnaire for all three surveys.

On July 1st, the detailed planning of the operation was provided from the directory, which considered seven working routes for visits to 126 sample locations. It was planned that the survey would take place between July 14th and August 10th, 2014. In order to include possible modifications to the scheduled dates, the CIDE coordinated meetings with CONAFE to jointly review the work plan and take those days which were not possible to go to any town into account, mainly due to the inability to locate respondents. Among the reasons that several homes in the same locality could not be located are the days government programs make payments (such as the Oportunidades Program), the festivals that people go to in neighboring towns to sell their products, or due to temporary fairs or markets that are on nearby or adjacent sites.

Additionally, the CIDE provided company field staff with a letter to assure institutional support during their visits to the villages. It is important to have such a statement given certain security conditions in some states, especially because they provide security in two ways: to survey takers when visiting locations where they are not identified, and to the respondents receiving strangers to give them important data.

As planned, field work began on July 14th and ended on August 10th, 2014. From an operational perspective, the company had a total of 28 interviewers, seven supervisors, two operational logistics coordinators and three external auditors. For operational purposes, seven teams were formed that included a supervisor and 4 interviewers, each of which was assigned a working path. The makeup of the group adhered to the protocol established in training, by having at least four members on each team so the survey could be done in pairs at every home.

3.1 Supervision

Monitoring is a standard procedure for administering the interview protocol in CIDE surveys. The objective is to ensure the quality and integrity of the information, particularly considering that these surveys would be conducted by persons unaffiliated to CIDE/BM to children and with a special type of interaction. It should be noted that during the surveys conducted for this project two-way monitoring is performed: by CIDE staff and by the company taken from its internal monitoring and information auditing.

Regarding the CIDE team supervision, unlike previous surveys, one of the most important specific objectives focused on the administration of ASQ questionnaires, without leaving aside the context questionnaires. In addition to this fundamental guideline, a direct supervision strategy was designed for conducting the ASQ that featured two main aspects: specialized traveling supervision by the psychologist who provided the training and the use of IT to approve the criteria established in the application of these tools.

In the first, specific routes were designed so the specialist was able to observe the interviewers. The design of the routes was based on part of the selection process in the training evaluation, such as those survey teams with average and low points in certain important aspects required for the proper administration of the ASQ tools. So these modifications could be applied and approved in real time for all survey takers, an electronic network of mobile phones was distributed among all CIDE supervisors in different states. The network not only allowed modification approval in real time, but also the ability to answer questions that eventually emerged and communicate them through the network to all supervisors.

With regard to the company, monitoring was mainly conducted in two ways: direct or coincidental supervision and subsequent supervision. Direct supervision was to provide support to each of the interviewers and to verify that surveys were conducted adequately, both locating correct addresses to be surveyed and proper application of the questionnaires. If a fault was detected, both the - company and CIDE supervisors – made the correction after the interview and followed the accompaniments until satisfied with the administration protocol of interviewers.

A company supervisor performed the subsequent monitoring by selecting the questionnaires from each of the interviewers at random and visiting households that had been surveyed. This way the supervisors were able to verify that interviewers had visited the household and by asking some questions, could confirm that the person listed as the application protocol had indeed been interviewed. Additionally, the supervisor routinely checked that the questionnaires had been correctly filled out, if not monitor directly in the field, before sending them to headquarters for data capture. To perform this review, the general guidelines for checking field research questionnaires were used.

3.2 Indigenous Languages

In the previous report the difficulties arising from the strategy of hiring translators who were from the communities visited were discussed; this mainly related to the translators who knew the language, but not the variant of the town visited. In this third survey, the same strategy that was used as in the second: hiring people *in situ* with knowledge of the local variants of the languages. Thus, efforts were coordinated directly with the town authorities, especially with the two leading figures in the operational chain: CONAFE promoters and supervisors. The primary objective was to continue to provide accuracy to the interpretation of the information gathered. The result was that local people, such as authorities or experienced interpreters were usually hired.

In all three stages, Oaxaca and Chiapas were the only states where interpreters were required. The protocol to use the tools with the aid of interpreters established the following flow: the fieldworker always faces the respondent and then the interpreter asks the question. Prior to the start of interviews with interpreters, the interviewer asks for consent by the respondent to record audio or video so that the quality of the interpretation of the answers could be validated at a later date. The Table below shows the languages and their variants in both states.

Table 3: Indigenous Languages and Dialects by State

Chiapas	Oaxaca
Tzotzil	Mixteco from the Northwest
Tzeltal • From the East • From the North • From the South	Mazateco • From the Presa • From the Southeast • From Eloxochitlan
Chol	Zapoteco from the Sierra

4. Capture and Validation

The polling company began the capture and validation process of the data collected in the first communities at its headquarters in August; this was the same process used for the baseline survey, while for the second survey the process was performed by a CIDE/BM team. For the company to start the process the CIDE/BM team provided the database design format for use with SPSS statistical software (* .sav). The aim of this measure was to standardize, as far as possible, the survey variables with those contained in the databases from the data captured by the company.

The company performed both processes (capture and data validation) using the CSPro package (The Census and Survey Processing System). The validation criteria were designed by the company and focused on verifying the consistency of the information contained in the questionnaires; a type of intelligent capture was used that allowed them to identify different categories of errors. This resulted in the delivery of 25 databases readable in SPSS. Ten of these databases are: household, caregiver, pregnant women, fathers, and CBQ; in five cases of treatment and control. The other database integrated information collected through ASQ tools for the last fifteen age ranges of 14 - 60 months.

5. Descriptive Statistics in Panel Surveys

The results shown below are intended to put the information from the three surveys conducted during this evaluation into a comparative perspective. Mainly context variables of households that were part of the sample are shown; however, it begins by looking at the results related to the response rate in the last two waves of the panel. Next, certain variables describing the sample households, their members and housing characteristics are shown. The section ends with the results of caregivers, pregnant women and fathers.

5.1 Home Response Rate

One fundamental point of interest regarding quality during the survey cycle is the total survey error that can occur and its impact on the results. The type of non-sampling errors, those that do not come from the sample design but are found in the errors of observation and lack of observation, are of particular concern in this report. In this section we focus on one of the integrated errors due to lack of observation that are usually of substantial consideration in the operation of a survey: non-response of primary observation units, which are households[8] in this study.

It must be mentioned that this study considers a relative response rate, defined as such due to a contextual factor of the study: households selected to participate came from a sampling of non-probability convenience through a sample unit rigging technique (propensity score matching), therefore, the basis on which the non-response rate is calculated is not a representative sample of households. It is also important to note that not all households where the household questionnaire was applied during the first stage met the primary criteria to continue in the study and was subsequently removed from the database. Note that this adjustment had an impact on the estimated response rate described in this section but was not performed for the variables included in the study.[9]

[8] The other subcomponent of non-observation errors is finding no response to individual items on an instrument, that is, questions that for various reasons have no answer marked on a questionnaire yet had been answered by the respondent.

[9] The number of households that were recorded in the baseline was 1386. However, all households that were identified as non viable for the project, after being analyzed, were removed based on the following criteria: not having children in the required age range and not having any member who was pregnant.

Households distributed in 126 communities were finally selected to join the project, these towns were located in six Mexican states: Chiapas, State of Mexico, Oaxaca, Puebla, Queretaro and Veracruz. In descending order, you can see that most communities were located in the State of Mexico for a total of 28.6%, next are the states of Queretaro and Veracruz with 20.6% each, Chiapas follows with 14.3% and finally the states of Oaxaca and Puebla with 7.9% each (Table 4).

Table 4: Communities by Federal Entity

Federal Entity	Locations	
	Frequency	%
Chiapas	18	14.3
State of Mexico	36	28.6
Oaxaca	10	7.9
Puebla	10	7.9
Queretaro	26	20.6
Veracruz	26	20.6
Total	126	100

The household response rate was generally higher in the third survey than the second by reaching 80.1%, 3.2% higher than 76.9% obtained in the second stage of the study. Oaxaca is the state where the highest response rate was identified, a remarkable situation because it is repeated in both stages: 87.8% in the second and 89.8% in the third. Next is Queretaro with 89.2%, but only for the third survey since it ranked second lowest with 75.3% of households in the second. Chiapas was ranked third by reaching 80.2% of households for the third stage. Veracruz, State of Mexico and Puebla, meanwhile occupied the last places with 79.4%, 76.2% and 70.6% of total households respectively (Table 5).

In three of the six states, Chiapas, Puebla and Veracruz, the response rate of the third survey was slightly lower than in the second with 0.6%, 5.9% and 3.5% fewer homes than in the last stage. However in the remaining states, Mexico, Queretaro and Oaxaca, the rates were higher showing increases for the last stage of 4.9%, 13.9% and 2% respectively (Table 5). The increase in Queretaro of just under fourteen points for the third survey is of particular interest. Now, as these three entities comprise 57.5% of total locations and 61% of all households, the overall response rate is higher for the third survey than the second.

Table 5: Relative Response Rate in Households by Federal Entity and Survey Waves

Federal Entity	Household					
	Frequency	Distribution %	Response Frequency		Relative Response Rate %	
	I		II	III	II	III
Chiapas	172	12.6	139	138	80.8	80.2
State of Mexico	508	37.4	362	387	71.3	76.2
Oaxaca	98	7.2	86	88	87.8	89.8
Puebla	102	7.5	78	72	76.5	70.6
Queretaro	223	16.4	168	199	75.3	89.2
Veracruz	257	18.9	213	204	82.9	79.4

Note: The Arabic numerals "I", "II" and "III" refer to the first, second and third surveys.

More specifically, we found that the main reasons why data could not be collected from households differed between the second and third stages. In the second stage, the absence of people was the predominant reason reaching 11% of the households, while in the last stage, was the refusals to give information, situation observed in 4.4% of households. The second reason most often found in the non-response rate for both stages was a change of address, reaching 2.5% of households in the second and 4.2% in the third. Codes such as “nobody home” or “temporarily away”, used to indicate absences from the home, totaled 2.1% in the second survey, a rate that was surpassed in the third survey by three percentage points (5.1% -2.1%). In the last stage 3.6% of households had migrated compared to virtually none in the previous stage (Table 6).

Table 6: Interview Results

Interview Result	Baseline		Second Survey		Third Survey	
	Frequency	%	Frequency	%	Frequency	%
Interview completed	1360	100	1046	76.9	1088	80.1
Interview not completed or interrupted			18	1.3	0	0
Refused participation			28	2.1	60	4.4
Person not home			150	11	14	1
Change of residence			34	2.5	57	4.2
Temporarily absent			23	1.7	35	2.6
Entire family migrated			1	0.1	49	3.6
Nobody home			5	0.4	34	2.5
Other			55	4	22	1.6
Total	1360	100	1360	100	1360	100

Diverse reasons (others) reached between 4% and 1.6% in the second and third stages respectively (Table 6). Among others, the reasons contained in this result are as follows: home in the sample not located, or by company supervisors or CONAFE promoters; for security reasons, it was not suitable to visit the home; there was unwillingness on the part of the members; merging households; and people who did not live in the home, but if they had indeed migrated could not be confirmed by neighbors. The fact that incomplete or interrupted declined interviews were reduced to zero is significant; this situation is supported through an agreement with the company that in the third survey they would classify the result as a survey taker's inability to conclude homes where a survey had been started. Derived from this survey takers were encouraged to finish, or reclassify the result more accurately and based on the protocol.

Regarding assessment location types, the treatment had a lower response rate than the control in both stages. Treatment communities reached a response rate of 73% and 78.7% in the second and third surveys, respectively, while the control 81% in the second and 81.4% in the third. The factors that most affect response rates showed variations depending on the stage and type of locality. Recall that in the second stage of the study the lack of people in general was the main reason for not responding in homes. However, there are differences between the two types of locations: in treatment communities this factor was six percentage points higher than in the control to reach 14.1% in the first, compared with 7.9% in the second. During the third stage, the absences were almost nil

while the two further factors that affected the response rate in treatment localities were migration of households at 5.4% and control for address changes at 4.8% (Table 7).

Table 7: Relative Response Rate in Households from Second and Third Survey According to Interview Results and Type of Community

Interview Results	Type of Locality					
	Baseline Frequency		Second Survey Percentage		Third Survey Percentage	
	T	C	T	C	T	C
Interview completed	688	672	73	81	78.7	81.4
Interview not completed or interrupted			2.3	0.3	0	0
Refused participation			1.7	2.4	4.9	3.9
Person not home			14.1	7.9	1	1
Change of residence			2.2	2.8	3.6	4.8
Temporarily absent			1.3	2.1	2.6	2.5
Entire family migrated			0.1	0	5.4	1.8
Nobody home			0.3	0.4	2	3
Other			4.9	3.1	2.6	1.6
Total	688	672	100	100	100	100
Note: "T" represents Treatment communities whereas "C" represents control.						

Within the scope of the experiment, it is important to note that the difference between the response rates of treatment and control communities dropped from the second to the third survey, with 8% in the second stage, and 2.7 % in the third. This situation presented a more favorable landscape for impact estimates between baseline data and the third survey since in the latter, groups are a slightly more balanced in terms of the response rate (Table 7).

The results by state highlight Oaxaca, where people were more participatory in the last two stages. Not only because in both surveys the highest response rate was shown in this entity, as previously noted, but because it is the only state in which none of the households that were visited refused to answer. About the same interview result (refusals) stresses that in states of Chiapas, Mexico, Puebla and Veracruz, non-response rate were higher from second to third survey, increasing by 4.1%, 4.1%, 3% and 0.4% respectively; Queretaro remained at the same level for the two stages -1.8% -. In fact this was one of the main factors in Puebla that explain the non-response rate (Table 8).

The state that recorded the lowest increase in refusals to provide information was Veracruz, unlike Chiapas, where the largest increase is observed. Part of what could explain this increase, at least in the states of Mexico and Chiapas, is the prevailing insecurity. In some places it was difficult for interviewers to convince inhabitants to participate, even when accompanied by CONAFE officials; in other cases there was no access to some households.

Table 8: Interview Results by Federal Entity

Interview Results	Chiapas		State of Mexico		Oaxaca		Puebla		Queretaro		Veracruz	
	II	III	II	III	II	III	II	III	II	III	II	III
Interview completed	80.8%	80.2%	71.3%	76.3%	87.8%	89.8%	76.5%	70.6%	75.3%	89.2%	82.9%	79.4%
Interview not completed or interrupted	0.6%	0%	2.8%	0%	1.0%	0%	0%	0%	0%	0%	0.8%	0%
Refused participation	2.9%	7%	1.6%	5.7%	0%	0%	2.9%	5.9%	1.8%	1.8%	3.1%	3.5%
Person not home	14.5%	0.6%	15.4%	1.4%	7.1%	1%	9.8%	1%	7.6%	1.3%	5.1%	0.4%
Change of residence	0%	1.7%	2.4%	6.3%	0%	1%	3.9%	4.9%	2.7%	2.2%	4.7%	4.3%
Temporarily absent	0%	4.7%	0.2%	2%	2%	2%	2%	5.9%	6.3%	1.8%	1.6%	1.9%
Entire family migrated	0%	2.9%	0%	3.9%	0%	4.1%	0%	3.9%	0.4%	0%	0%	6.2%
Nobody home	0%	2.3%	0.2%	1.8%	0%	2%	1%	4.9%	0.4%	1.8%	0.8%	3.9%
Other	1.2%	0.6%	6.3%	2.6%	2%	0%	3.9%	2.9%	5.4%	1.7%	1.2%	0.4%
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Note: The Arabic numerals "I", "II" and "III" refer to the first, second and third surveys

The third survey data also indicates that change of address was the reason for a higher non-response rates in the state of Mexico and Queretaro (6.3% and 2.2%), while in Chiapas they were refusal to respond (7%). In Puebla refusal to respond (5.9%) and temporary absences (5.9%) were the same. Migration of households prevailed in Veracruz and Oaxaca occurring in 6.2% and 4.1% of households respectively. The second factor most often mentioned in Chiapas was the absence of people (4.7%), in the state of Mexico refusal (5.7%) and Veracruz changes of address (4.3%). Temporary and all absences from the home were the second factor most often mentioned and with the same ratio factors in Oaxaca and Querétaro: 2% for the first state and 1.8% for the second. In Veracruz, meanwhile, the second interview results were more often a change of address with 4.3% (Table 8).

In short, the non-response rate was mainly concentrated in the states of Mexico and Puebla for both second and third surveys. While the rate was not entirely favorable in Queretaro for the second stage, the recovery of households for the third is notable. Chiapas, Oaxaca and Veracruz were relatively stable throughout the study. Finally, it was the treatment communities that reported a lower response rate in two stages. However, their recovery stands out for the third phase of data gathering, which is of substantial importance in relation to the evaluation method use in this research.

5.2 Attrition on Caregivers and Children

As previously mentioned there was a diversity of questionnaires applied through the impact assessment. For this reason there are two levels in the non-response rate: the household level described above and the individual one that mainly focuses on the rates at which caregivers and children independently participated on the survey. Therefore, at this point it is of importance to note that individual response rates may be independent of that at the household level, but not conversely (id. est. whenever a household could not be reached then none of its members were surveyed, but when a household was surveyed did not mean that all of its members participated). Made the clarification, then the non-response rate in caregivers and children as well as their effect on the impact estimates are presented in this research.

In this evaluation, some proportion of children and caregivers attrited from the program. As Table 9 below shows, attrition was relatively high between the baseline and Wave 2, at which point 69% of children in treatment were re-tested and 76% of children in control, with similar numbers for caregivers. Wave 3 shows an improvement both in overall numbers, where 78% of treatment children and 81% of control children were re-tested, with slightly higher attrition among caregivers.

Table 9: Attrition Descriptive Analysis

	Children		**Caregivers**	
	Treatment	Control	Treatment	Control
Baseline	797	788	690	672
Wave 2	553	598	472	493
	(69%)	(76%)	(68%)	(73%)
Wave 3	623	638	505	523
	(78%)	(81%)	(73%)	(78%)

The principal concern with attrition is that it may be unbalanced. If attrition is particularly stark among the comparison group (or the treatment group), for example, then estimates of impact may be biased. We test for unbalanced attrition between treatment and control using linear probability models with an indicator variable for having attrited in Wave 2 (or, in another regression, in Wave 3) regressed against a dummy for assignment to treatment, based on community of residence. As in all of our regressions, we cluster standard errors at the community level, as that is the level of random assignment. As Table 10 below demonstrates, while attrition is slightly higher among the treatment group, the difference is not statistically significant. Given the lack of significant differences between the groups, we make the reasonable assumption that attrition is not driving impact estimates.

Table 10: Attrition Regression Analysis

	Treatment	Control	P-value of difference
Children			
Wave 2	0.31	0.24	0.11
Wave 3	0.29	0.25	0.25
Caregivers			
Wave 2	0.32	0.27	0.20
Wave 3	0.27	0.22	0.15

5.3 Home

An important fact regarding quality of information in a longitudinal study is whether the respondents in various survey waves, are people who have no notion of their participation in terms of experience as a respondent. This provides consistency to the data collected since the same people who already participated in the study are selected again. In this regard, no substantial changes were observed as to whether the respondent had been in previous surveys or not; the difference between the second and third stages was approximately 2.7%. However, a positive change in the latest survey is that 96% of respondents were also in one of the first two surveys; this compared to the second survey in which 93% of people who answered the questionnaire also participated in the baseline (Table 11).

Table 11: Is the person being interviewed the same person in this household that was interviewed during the previous survey done by CIDE?

Percentage by Survey		
	II	III
Yes	93.3	96
No	6.7	4
Total	100	100

Households on average recorded five members in all three surveys with slight variations in each: 5 in the baseline, 5.3 in the second and 5.6 in the third. The number of children under five years of age, on the other hand, changed during the two-year duration of the stages. At baseline approximately 80% of households had only one child under five, 18% with two and just over 1% with three or more. In the second survey, the proportion of households with children under five years of age decreased to 62.2%, while households with two children reached 32%; 5.5% of households had more than 3 children in that age range (Table 12).

Table 12: Children under 60 months of age in the household

Children under 5 years of age	Baseline	Second Survey	Children under 6 years of age	Third Survey
	Percentage	Percentage		Percentage
1	80.9	62.2	1	51.7
2	18	32.3	2	37.2
3	1	4.8	3	9.5
4 - 6	0.2	0.7	4 - 6	1.6
Total	100	100		100

Consistent with the logic of advancing age of members on a panel, you can see that the number of households with more than one child significantly increased from the first to the second stages, moving from 18% to 32%. Currently, the data indicate that almost 52% of households have only one six-year-old child, fewer than 37% have two and 9.5% have three; 1.6% between 4 and 6 children in the same age range (Table 12).

Table 13: LANGUAGES Does [….] speak Spanish? Members 5 years of age or older.

	Percentage		
	Baseline	Second Survey	Third Survey
Yes	93.4	93.6	95.5
No	6	5.9	4.3
Cannot	0.1	0.1	0
Unknown /no response	0.5	0.3	0.1
Total	100	100	100

One important feature among the target population is the languages they can speak and write. In this regard we find that the percentage of Spanish speaking household members five years of age or over presented an increase during the study, especially in the third year. In baseline 93.4% reported speaking Spanish, 93.6% in the second survey and in the third, 95.5% (Table 13). Contrarily, the proportion of members aged 10 and older who can read and write a message, either in Spanish, an indigenous language or in both decreased. The percentage of individuals who can read and write a message in Spanish only, decrease from a baseline of 85.7% to 85.6% and 85.2% in the last two stages (Table 14).

Table 14: Does [….] know how to read and write a message? Members 10 years of age or older.

	Percentage		
	Baseline	Second Survey	Third Survey
Yes, only in Spanish	85.7	85.6	85.2
Yes, only in an Indigenous Language	2	1.6	0.8
Yes, in Spanish and an Indigenous Language	4.2	7.3	5.9
No	4.9	4.8	7.4
Cannot read/ write	0.6	0.6	0.5
Unknown	2.6	0	0
Total	100	100	100

With levels of education for household members aged 17 or older, we see that the percentage of those who have no formal education was lower in the third survey than in the first two; 5.8% was reached in the current survey while 6.4% was recorded in the baseline and 6.8% in the second. There was a high concentration of basic education level for household members during all three stages, registering 78.7%, 80% and 77.1% in the first, second and third respectively. Meanwhile, in the senior high school level of education

showed proportions of 11.4%, 9.2% and 11.7% reported from the first to the third survey. The percentage of household members with a university degree went from 2.4% to 0.9% then 1.9%, in the first, second and third stages, in the same order (Table 15).

Table 15: Highest level of education completed regarding members 17 years of age or older

	Percentage		
	Baseline	Second Survey	Third Survey
No formal education	6.4	6.8	5.8
Preschool	1.1	1	0.8
Primary	42.9	47	40.7
Junior High	35.8	33	36.4
Senior High	11.4	9.2	11.7
University Degree	2.4	0.9	1.9
Total	100	100	100

Note: responses do not apply, does not know and no response equal a total of 2.1% and 2.7% for the second and third stages.

The housing conditions varied in the same proportions during the three surveys. On average three rooms were recorded in total, including the kitchen; 69%, 68% and 66% of households reported having three rooms on the first, second and third stages respectively. The average number of rooms used for sleeping ranged from 1.7 and 1.8 in the three surveys, while the proportion of households with one or two rooms for the same purpose was 84% at baseline, 85% in the second stage and 83% in the third.

Table 16: Of what material is the majority of the floor in this house made?

	Percentage		
	Baseline	Second Survey	Third Survey
Dirt	9.1	8.9	7.8
Cement or Solid	79.8	80.3	80.7
Other material (wood, tile or other)	11.1	10.8	11.5
Total	100	100	100

With regard to construction materials of the houses, although with some variants, the results over the three years do not seem to have been very substantial. The proportion of houses with dirt floors is lower as one moves through the three stages of the study, 9.1% at baseline, 8.9% in the second and 7.8% in the third floor. The trend was the opposite

however for households that reported having cement or solid floor, where at baseline 79.8% of households here in this condition, in the second stage there were 80.3% and in the third 80.7%. Meanwhile, 11.1% of households at baseline mentioned another surface, 10.8% in the second stage and 11.5% in the last (Table 16).

Table 17: Of what material is the majority of the walls of this house made?

	Percentage		
	Baseline	Second Survey	Third Survey
Waste material	0.3	0.1	0
Cardboard sheets	0.4	1.3	0.1
Tin or asbestos sheets	1.8	0.3	1.6
Reeds, bamboo or mud covered sugarcane	0.4	0.5	0.7
Mud or mud covered sugarcane	0.5	0	0.7
Wood	13.2	12.6	13.6
Adobe	6.6	8.7	8.1
Brick	76.9	76.6	75.2
Total	100	100	100

We found most of the walls were made of brick in all three stages with about three-quarters of all households with using this material. Wood is the second most common material for walls as households reported a slight increase in share from 13.2% in the first to 13.6% in the third survey. Adobe is the third most often used material for walls, reaching 6.6% at baseline, 8.7% in the following year and 8.1% in the final. Finally, walls made of other materials, including waste material or cardboard sheet ranged between 2.2% and 3.4% over the three phases (Table 17).

Table 18: Of what material is the majority of the roof of this house made?

	Percentage		
	Baseline	Second Survey	Third Survey
Cardboard sheets	4.5	4	3.3
Tin sheets	32.5	32.1	35
Asbestos sheets	7.8	9	7.6
Wood	0.7	0.7	0.6
Clay tiles	2.8	2.4	2.8
Concrete slab	51.5	51.8	50.8
Waste materials	0.1	0	0.1
Palm leaf or straw	0.1	0	0
Total	100	100	100

Roofing materials also have very similar conditions throughout all phases of the study. Just over half of all households reported that most of the roofs of their homes are built with concrete slab, 51.5%, 51.8% and 50.8%, in ascending chronological order. Tin is the next most common material used for the same purpose, which employs 32.5% reported for dwellings in baseline, 32.1% in the second stage and 35% the third. Asbestos sheets, cardboard and clay tiles are the roofing materials used in 11.5% and 13.7% of all households. The use of waste material and palm leaves, although very low or zero in the three stages, were indeed recordable data, (Table 18).

Table 19: During the last week, how many days was power available for at least three hours per day?

	Percentage		
Days	Baseline	Second Survey	Third Survey
0	3.5	4	2.8
1	0.5	0.3	0.5
2	0.5	0.6	0.3
3	0.6	1.1	1.8
4	0.3	1.4	2
5	1	1.5	2.5
6	1.9	3	2.6
7	91.8	88.2	87.5
Total	100	100	100

Access to electricity was measured by asking respondents in how many days was there electrical power for at least three hours in a week. As a result we note that the average number of days at baseline was 6.9, in the second survey of 6.5 days and the third of 6.8 days. From the frequency perspective, between 2.8% and 4% of households continue to report not having electricity any day of the week. Worth mentioning is that during visits, although several homes were identified as being within locations, the distances between them was quite significant and the roads quite rough. In these locations, basic services like water and electricity are extremely scarce (Table 19).

In any case, a large portion of households reported having electricity service throughout the week, 91.8%, 88.2% and 87.5% for baseline, second and third stages, respectively. For Baseline, the percentage of households that stated having electrical power just one to three days was 1.6%, in the second stage it was 2% and 2.6% in the third. What stands out is that the percentages are higher if we consider households that have electrical service from four to six days, 3.2%, 6.9% and 7.1% in the same chronological order. Therefore, it is generally observed that approximately 95% of all households in the sample have service at least four days a week in all three surveys; the remaining households seriously lack such service and, as such, the possibility of many other resources and related services (Table 19).

Table 20: To which type of drainage or sewer systems is this house connected?

	Percentage		
	Baseline	Second Survey	Third Survey
Public system	33.4	32.3	32.4
Septic tank	33.6	32.3	35.9
Pipe to a drop off	4.2	5.5	5
Pipe to a river, lake or sea	3.5	2.8	2.5
No drainage	25.3	27.1	24.3
Total	100	100	100

Sanitary conditions of households are relevant for further knowledge about health and thus an important variable for the development of children. In this sense, we see that over the panel a considerable percentage of households continued without drainage and thus lack favorable health conditions; the percentages in the three surveys ranged between 24% and 27%. Connection to the public sewage system and use of septic tanks each covered approximately one third of households in all three surveys. Meanwhile, the proportion of households whose pipes flow into ravines or a body of water were found to range from 7.5% and 8.3% in the three stages. The situation in this area is somewhat unfavorable, particularly for children, since according to data collected there are still many households in areas considered to be hazardous to health (Table 20).

The availability of appliances in the home is a rough indicator of the decline in domestic work loads, as well as the use of time, comfort, personal hygiene and good nutrition of household members. In this category you can see in all three surveys that between 66% and 68% of households reported having a gas stove, while wood stoves were reported in 58%, 69% and 70% at baseline, second and third stages respectively (Table 21).

With respect to personal hygiene, it was noticed that most households reported having a water tank in the three stages. We also found that between 23% and 25% of households mentioned having a shower for all stages, a situation worth noting because about three quarters of households do not have facilities for personal hygiene and therefore families are living under unfavorable conditions for welfare and development. Moreover, the power consumption meter installation presented higher recurrence reaching rates of 79.3%, 79% and 82.4% in each of the surveys, in chronological order. The data initially shows certain consistency with the proportions of households who mentioned having the electrical service. However, this also demonstrates the opposite side of the panel, where between 13% and 15% of households had informal electricity service (Table 21).

Table 21: Appliances, goods and services in the homes.

	Has…(%)		
	Baseline	Second Survey	Third Survey
Gas stove	68.7	67.3	66.5
Wood stove	58.2	69.6	70.1
Water tank	47.8	50	48.6
Boiler	19.4	17.1	16.3
Cistern	13.9	15.7	16.2
Shower	25.3	23	23.6
Power meter	79.3	78.9	82.4
Automobile	24.2	23.8	20.7
Internet	2.8	1.7	2.8
Computer	5.2	3.7	6.5
Telephone	10	8.4	7.6
Cellular	50.3	53.2	58.6
Washer	33.5	33.1	33.1
Refrigerator	53.2	51.8	54.8
Television	84.5	82.9	84.1
Radio	61.7	58.1	56.6

An automobile serves as an indicator in the decrease of transportation time, a resource that is of particular importance considering the location of the villages, the roads conditions and the treatment communities and the potential need for people to come to early education meetings in remote areas. In this regard we note that in all three surveys, the proportion of households not owning a car was very high, ranging from 76% to 80% (Table 21)

Moreover, the capacity for safe food storage, nutrition and hygiene are closely related to goods such as refrigerators and washing machines. In the three stages of the study the majority of households reported having refrigerator; 53% at baseline, 51.8% and 54.8% in the last two surveys. While a washing machine was identified as an appliance which is found in only about a third of the homes during all stages of the project. Given these results throughout the project, between 47% and 48.2% of households did not have a basic appliance to support a proper diet and about 65% did not have support for improved hygiene in the household (Table 21).

The presence of and access to electronic devices and services that enable data processing, signal reception and in particular long distance communication is crucial at present; this also serves as an economic development indicator in the conditions of households. As for the electronic devices, from the first instance and throughout all stages, a television was reported in many homes, reaching between 82.9% and 84.5% of total households. This was followed by the radio, an article found in just over 61% at baseline, 58% in the second year and 56.6% in the third. A computer was found to be far from these proportions as it reported in just 5.2%, 3.7% and 2.8% of households in the first, second and third stages respectively (Table 21).

Now, although not all such services are recurring in homes, at least one quite accessible today was reported: the cell phone. Most homes have mentioned this device on all surveys, 50.3% at baseline, 53.2% in the second survey and 58.6% in the third. However, internet service was found to be much lower, in terms of other variables previously analyzed. In the baseline and the third stage only 2.8% of households have this service while just 1.7% in the second. The landline, moreover, was found in just 10% of households in baseline, 8.4% in the second stage and 7.6% in the third (Table 21).

According to the results presented regarding household conditions and their dwellings, we see that the proportions have not changed dramatically over the implementation of the housing information questionnaires. The description of the variables denote that the general conditions of households and dwellings in each of the surveys, independent of the statistics, remain unfavorable for a good part of the target population. Educational attainment, housing characteristics and the described economic indicators suggest that the public interventions that, according to this study, have intended to serve populations in these conditions should be specially designed and implemented cost-effectively to achieve their objectives and ultimately have a positive impact on development.

5.4 Caregivers

Primary caregivers are the most important figure in the early education intervention model. These caregivers are defined as persons, member or not of the household, that spend time raising the children in the household, and assume responsibility for their upbringing and make decisions in this regard. The figure of caregiver represents a fundamental link in the operational chain of program and the results flow. Progress in the operational chain is through the caregiver, given that their knowledge about early childhood education (pre-

school model), and as institutional transmitters of that knowledge (promoters) and its application to children, is connected. Product flow is included because this figure is part of targeting intervention and therefore a necessary subject in the impact assessment.

A pre-note describing the results, and also something emphasized in the report of the second survey, is about how to integrate the database of main caregiver descriptions are estimated. Caregiver databases from the surveys are relatively independent, this means that integrated in the bases of the second and third surveys, caregivers are not necessarily exactly the same respondents in the baseline for two main reasons.

The first is due to a natural change in the conditions of the context in which the primary caretakers, for various reasons, were interviewed at baseline, but are now found in the non-response rate group. In these cases, for example, mothers can be found who needed to work and let a family member or close friend serve as primary caregiver to their children. The second situation is the integration of new caregivers who support in the work of raising children and, in the case of the treatment group, also participate in the early education sessions. In short, the bases do not necessarily have the same caregivers, either because they were replaced or because new ones were added throughout the study.

As stated above, it is important to report that 88% of caregivers interviewed in baseline were the same number as in the second and the third surveys; nonetheless, even though the proportion is the same, it should not be assumed that they were exactly the same respondents at different stages. During the third survey, about 8.5% of caregivers were not at baseline compared with 12% in the second year; the remaining 2% in the third stage were new caregivers who had never been interviewed. Consistent with longitudinal study characteristics, caregivers have to have lived in their communities for 18, 19 and 21 years for baseline, second and third stages, respectively.

Table 22: What relationship do you have with [….]?

	Percentage		
	Baseline	Second Survey	Third Survey
Mother	91	90.6	90
Father	0.7	0.9	0.8
Brother	0.1	0.5	0.5
Grandparent	5.5	6.4	7.1
Other	2	1.6	1.5
Total	100	100	100

As for the relationship between the caregiver and infant(s) cared for, the proportions have not varied substantially between surveys. Not surprisingly, during the three surveys, it is the mother who cares for the children in approximately 90% of cases. Similarly, in the three stages, grandparents represented the second figure in homes that served as caregivers, reaching 5.5% in the baseline, with 6.4% and 7.1% in the second and third stages. A reasonably low percentage of fathers as caregivers was recorded for the three stages with percentages ranging between 0.7% and 0.9%. Other relatives and associates individuals accounted for approximately 2% of the caregivers of children at all stages (Table 22).

The results of the second survey found that caregivers had average of two and a half years caring for children while in the third it was three years, four months. Now, although these figures are by definition represent those that spend most of the time with the children, a good proportion of them reported to have one or more persons helping them in these activities. At baseline 73% of caregivers they reported having someone else helping them with the infants in the first follow-up 76% and approximately 68% in the last.

Table 23: Is someone in [….]'s household currently a recipient of the Oportunidades Program?

	Percentage		
	Baseline	Second Survey	Third Survey
No	55.5	49.8	45
Yes	44.5	50.2	55
Total	100	100	100

Government support is crucial to the quality of life of households with respect to food, health and economy. Despite the shortcomings that permeate these communities, we found that 55.5% of caregivers reported not being a recipient of Oportunidades in baseline, 49.8% in the second year and 45% in the third. According to the data, the trend could in fact increase the frequency of households who do receive such support (Table 23).

The Oportunidades program, while an important intervention, others also exist that serve different primary needs for this population. Programs, such as Seguro Popular, are dedicated to reducing catastrophic health expenditures potentially incurred by individuals who lack health insurance. It was found that the proportion of caregivers who mentioned that there is at least one member of the Seguro Popular program in the household, has increased by just under three fifths at baseline (59%), to almost three quarters (74%) in the last stage (Table 24).

Table 24: In [….]'s household, has anyone received support from: Seguro Popular, Procampo, Scholarships or educational programs for senior citizens?

	Percentage		
	Baseline	Second Survey	Third Survey
Seguro Popular	59.4	68.7	74.3
Procampo	6	5.3	5.6
Scholarships	9	11.4	13.6
Seniors programs	4.6	5	5.7

Similar statistical behavior was identified in educational support programs for the elderly, which increased in each of the surveys. Regarding the first type, throughout the study between 9% and 13.6% of caregivers mentioned that in their homes they have received support through educational activities. Meanwhile support for seniors, to a lesser extent, was also growing in proportion going from 4.6% at baseline, 5% in the second and 5.7% in the third survey (Table 24).

Table 25: In [….]'s household, has anyone received support (such as cash, remittences, clothing, foodstuffs, etc.) from people that live in another country or in Mexico during the last year?

	Percentage		
	Baseline	Second Survey	Third Survey
Financial support from people living within the country	6.2	4.7	4.5
Financial support from people living in an other country	3.3	4.5	5.7

Other support mechanisms were also inquired about, in particular, economic or similar and those from nongovernmental sources. In this category frequency was negligible with caregivers mentioning these being; only 6.5% of caregivers said that home economic support from people living in the country, 4.7% in the second survey and 4.5% in the third was received; at each stage the proportion was smaller.

Notably the contrary occurred regarding support from other countries which caregivers initially reported at 3.3%, 4.5% in the second stage and culminated with 5.7% in the last (Table 25).

Table 26: Approximately how many books does [….] have?

	Percentage					
	0 to 35 months			36 to 60/72 months		
	I	II	III	I	II	III
Ten or more	3	6.3	6.5	4.6	6.6	7.9
Three to Nine	13.3	23.2	24.7	20.4	27.6	32.6
One or Two	15.1	20.1	22	29.1	27.4	24
None	68.6	50.4	46.7	45.9	38.4	35.4
Total	100	100	100	100	100	100

Note: The Arabic numerals "I", "II" and "III" refer to the first, second and third surveys

Child development is closely related to parenting practices of the caregiver, thus a holistic concept represented in two sections of the caregiver questionnaire were according to two age ranges of children: 0-35 months and 36 months to 72 months. Educational resources, among others, represent substantial support in caregiver practices. It is observed that as the survey stages progressed, there was a decline in caregivers who mentioned not having children's books in both age groups. Contrary to this trend, the percentages of caregivers who reported having three or more books, increased in both age groups. Meanwhile, caregivers with one or two books followed the same trend during the stages of study but only for the range of 0-35 months, this was in contrast to the range of 36-72 months that reported consecutive declines from first to third surveys (Table 26).

Complementary to the possession of books, the frequency with which caregivers read or tell stories to children was asked about. In this area the data indicates that the percentages of caregivers who failed to conduct this activity were decreasing in both age ranges throughout the panel. In the first range, the percentage of caregivers in this

situation dropped from almost half -49.7% - at baseline to nearly a quarter -24% - in the last stages of the study, while in the older range, from about one third -33.7% - to a fifth -20.5% - (Table 27).

Table 27: How often you have the opportunity to read or tell stories to [….]?

	Percentage					
	0 to 35 months			0 to 35 months		
	I	II	III	I	II	III
Never	49.7	32.6	24	33.7	20.9	20.5
A few times per year	3.4	7.1	2.4	3.6	8.9	5.7
A few times per month	8.2	11.3	16.2	18.7	16.7	17.9
Once per week	15.8	20.3	27.5	22.8	25.4	28.3
Three times per week	11	15.8	18.6	11.4	17.1	16.9
Everyday	11.6	13	11.3	9.8	11.1	10.7
Total	100	100	100	100	100	100

Note: The Arabic numerals "I", "II" and "III" refer to the first, second and third surveys

Contrarily, there was a declining trend between the extreme phases of the study: the proportion of caregivers who reported reading to children every day was less between the first and third surveys for both age ranges, although with a relative increases in the middle stage. A similar trend was observed in caregivers who mentioned reading a few times a year. In contrast is the propensity of those who said they read once a week, as the percentages of caregivers in this situation were increasing for both age ranges and for each study phase.

Meanwhile, in the range of 0-35 months, the proportion of caregivers that mentioned reading a few times a month or three times a week consistently increased from 8.2% and 11% at baseline to 16.2% and 18.6% in the last survey. The situation was different for caregivers who reported the same in the age range of 36-72 months, where the proportions decreased from 18.7% to 11.4% at baseline to 17.9% and 16.9% in the third stage (Table 27) .

Table 28: Some parents dedicate time to teaching their children, while other parents believe that children learn better on their own. Which of the statements below best describe what you think?

	Percentage		
	Baseline	Second Survey	Third Survey
Parents should always spend time teaching their children	67.2	65.1	64.6
Parents should always leave their children to learn on their own	7.4	4.5	10.3
Parents should usually spend time teaching their children	21.8	28.6	22.3
Parents should usually leave their children to learn on their own	3.6	1.9	2.6
Total	100	100	100

As for the attention a caregiver should pay to the child, only in the 0-35 months range, the results between the baseline and the third survey indicate a decrease in the proportion of those who believe that time should be spent teaching their children, going from almost 90% in the first survey to 86.9% at the end of the study. Conversely, the results at baseline of caregivers who believe that parents should let children learn on their own reached 11% while increasing 12.9% in the third (Table 28).

With regards to the use of physical punishment on children we found the following. The proportion of caregivers of children 0-35 months that never applied such measures decreased from 65% at baseline to 57% and 47% in the second and third surveys. You can also see that the greatest percentage difference corresponds to caregivers of children in the range 0 to 35 months that said they never applied any physical punishment but now they do, which reached almost 20%. However this trend was reversed in the 36-72 months range with an increase of nearly 48% of caregivers at baseline to fully half in the third stage, though with a greater range during the second survey -51.9% - (Table 29)

Table 29: Sometimes children are well behaved and sometimes not. During the last week, approximately how many times, if it occurred at all, did you have to spank [….]?

	Percentage					
	0 to 35 months			36 to 60/72 months		
	I	II	III	I	II	III
None	65.1	57.56	47.4	47.9	51.9	50
One	17	19.8	23	19.8	24	24.5
Two	10.3	14.2	17.5	20.8	13.1	16.1
Three or more	7.5	8.48	12	11.5	10.9	10.3
Total	100	100	100	100	100	100

Caregivers who use physical punishment mentioned only once increased their percentages in each of the surveys for both age groups. In the lower age range the trend went from 17% at baseline to 19.8% and 23% in the second and third stages, while the greatest was from 19.8% to 24% and 24.5% in the same order. Those people that said they used these measures twice increased throughout the study, 10.3%, 14.2% and 17.5% in chronological order of the stages. The trend however was not the same in the older range as it began with 20.8%, then 13.1% and culminated at 16.1% (Table 29).

Among other parenting practices researched, were also those related to health prevention, entertainment and domestic activities, and this time both the primary caregivers and the fathers of the children were interviewed, when the father was not the caregiver. In the first category we see that in the baseline survey, 52% of caregivers mentioned taking children to a preventive medical examination, 58.6% in the second survey and 52% again in the third; these data suggest no substantial changes between the beginning and the end of the study. Entertainment reports between the caregiver and children also showed no substantial changes; 95% of caregivers claimed having played with the children during the baseline, which remained essentially the same in the second, but third survey recorded a slight drop to 94.2%. Singing to children by Caregivers was mentioned by 77% at baseline, 82% and 81.4% in the second and third stages respectively (Table 30).

Table 30: I am now going to ask you to please tell me if the following activities or situations occur:

	Percentage		
	Baseline	Second Survey	Third Survey
Do you take [….] for a medical check up even when not sick?	52.3	58.6	52.4
Do you play with [….]?	95.2	95.5	94.2
Do you sing with [….]?	77	82	81.4
Does [….] have a musical instrument, either purchased or hand made?	35.7	40.7	37.1
Does the father of [….] help prepare [….]'s meals?	44.2	43.6	47.5
Does the father of [….] help dress [….]?	67.4	67.9	69.4
Does the father of [….] play with [….]?	88	87.8	91.2
Does the father of [….] carry [….]?	90	88.4	91.9

Moreover, between the baseline and the last stage a relative increase in the proportion of caregivers commented that fathers participated in some childcare activities. At baseline 44.2% of caregivers said that fathers helped prepare the food, in the third 47.5%. Also, 67.4% mentioned in the first stage was that fathers help children get dressed, and 69.4% in the third. This trend was repeated in the activities of carrying and playing and with the children. During the baseline, 88% of the caregivers said fathers played with their children, in the final survey it were 91.2%. In the initial stage 90% reported that fathers held their children while in the end it was 91.9% (Table 30).

Table 31: Up to which level of education do you believe [….] will study?

	Percentage		
	Baseline	Second Survey	Third Survey
Primary	4.7	3.7	4.4
Junior High	25.9	22.2	20.7
Senior High	28.6	32.8	30.4
Degree	39.3	39.1	40.3
Postgraduate	1.5	2.2	4.3
Total	100	100	100

One last point refers to the expectations of educational achievement that caregivers have with regard to children. In this area the results indicate that there was an increase in the proportion of caregivers who expected to see children attend a higher education institute between the first and last phases of the study. At baseline, 40.8% believed that their children would obtain a university or postgraduate degree. This same belief reached 44.6% in the last phase of data collection. To a lesser extent, a similar trend was observed in those who believe that their children would get a secondary school education, given that in the first stage 28.6% was reported, while in the final 30.4%. Conversely the belief that their children would achieve some level of basic education is observed; baseline in this belief was 30% of caregivers while in the last stage it was 25% (Table 31).

5.5 Pregnant Women

In the second and third surveys the number of pregnant women who answered the questionnaire was significantly reduced because to stay on the panel it would have been necessary to be pregnant during at least two of the three surveys in order to have data that estimate changes in practices during pregnancy. On this line we see that in the first and second stages of data collection, 6.6% then nearly 8% of households, respectively, reported having one or more pregnant women, while on the third visit just 5% (Table 32). Of the women who responded to the third survey, 42% did so at baseline, 55% on the second visit and just 2% percent responded in both.

Table 32: Number of pregnant women in the home

	Percentage		
Number of pregnant women	Baseline	Second Survey	Third Survey
0	93.4	91.78	94.55
1	6.5	7.92	5.27
2	1	0.3	0.18
Total	100	100	100

Among others, the subject of preventive health was an important fact, which was measured by the number of medical check ups during their pregnancy. We found that 4% of respondents at baseline had had no medical check ups; for the second survey this was reduced to 3.8% and the third dropped to zero. The trend in the average pregnant woman getting medical check ups rose as the study progressed. On average pregnant women

reported 4.5 medical check ups during the baseline that increased to 6.8 visits a year later and for the latest survey it was 6.9. Respondents were also asked if the child's father had attended a medical check up during pregnancy, demonstrating a baseline report of 65%; the second survey ratio reached 70% and 72% in the third (Table 33).

Table 33: During pregnancy, did the father of the child accompany you to a medical exam?

	Percentage		
	Baseline	Second Survey	Third Survey
Yes	65.2	69.8	72
No	33	30.2	28
Total	100	100	100

Conversely, there is the issue risky behavior at this stage, which was captured by certain questions in the survey. A step in this direction is pregnant women taking medications or remedies that had not been prescribed by a doctor, as did 11.3% of pregnant women during the first stage of the study, 7.8% in the middle stage and 12.5% in the end (Table 34).

Table 34: During pregnancy, did you take any medicine or remedy that was not prescribed by a physician?

	Percentage		
	Baseline	Second Survey	Third Survey
Yes	11.3	7.8	12.5
No	88.7	92.2	87.5
Total	100	100	100

Another important section of the subject is the consumption of substances potentially hazardous to health. Smoking and drinking habits certainly present a great threat especially for abortion during the first trimester of pregnancy. Among these, alcohol and tobacco are two of the products that present a great health hazard. In this regard we see that in the baseline, 2% of mothers reported having consumed alcohol in the week prior to the interview, this ratio remained constant in the second survey and rose to 6.2% in the third. Tobacco was consumed by 1% of pregnant women during the baseline and was reduced to zero within the two surveys (Table 35).

Table 35: Over the last 7 days, how often did you …?

	Percentage		
	Baseline	Second Survey	Third Survey
…consume alcoholic beverages	2	2	6.2
…smoke	1	0	0
…drink coffee	60	65	69
…drink soda	72	82	82
Total	100	100	100

Of all risk products coffee consumption was recorded at the highest levels during the three stages of the study, although it is clear that this product presents fewer risks during pregnancy. The consumption of coffee had an upward trend throughout the study, 60% of pregnant women reported consuming it on the first visit, the proportion rose to 65% in the second and again increased to almost 69% in third. Another risk captured was consumption of soft drinks, in which 72% of respondents in the first survey reported having drunk them, with the proportion increasing to 82% in the second and remained constant for the third (Table 35).

Table 36: Is the food you are currently eating different from what you were eating before getting pregnant?

	Percentage		
	Baseline	Second Survey	Third Survey
Yes	27	30	33
No	73	70	67
Total	100	100	100

In this late stage of the study, poor eating habits are decreasing compared to previous surveys. At baseline 73% of those interviewed claimed they that had not changed their diet during pregnancy, while in the second survey the proportion was 70% and the third to 67% (Table 36). In addition to this question, we have 25.4% reported having eaten less at the first visit, which declined slightly in the second to 22% and increased to 37% in the third. The proportion of pregnant women who claimed to eat the same was 27.2% at baseline, 48% in the second survey and 40% in the third. Finally, a declining trend identified was

women who mentioned having eaten more during pregnancy. At baseline 52% said they eat more, 30% said the same on the second visit and 22% in the last (Table 37).

Table 37: Compared to what you were eating prior to the pregnancy, you are now eating ….

	Percentage		
	Baseline	Second Survey	Third Survey
…less	25.4	22	37
…the same	27.2	48	40.7
...more	52	30	22.2
Total	100	100	100

In terms of parenting practices, mothers were asked if they believe that children are aware of was going on around them or not. In the first two stages, 91% of respondents felt that babies are in fact aware of what is happening around them, whereas in the last survey the proportion dropped to 86% (Table 38).

Table 38: Do you believe infants are aware of what is going on around them?

	Percentage		
	Baseline	Second Survey	Third Survey
Yes	91.3	91.5	86
No	5.2	6.8	10
Total	100	100	100

Note: Does not know and no response make up the difference of the total

With regard to the expectations of educational achievement, a substantial change occurred in the proportions by stage. At baseline 28.7% of pregnant women believed their children would reach some level of basic education, while in the second survey 36.8% believed the same, however in the third it was just 22.6%. Approximately 30% of pregnant women in baseline believed that their children would achieve senior high school education, a proportion that grew in the following two surveys to 38.8% in the second and 41.9% in last (Table 39).

Table 39: Up to which level of education do you believe your child will study?

	Percentage		
	Baseline	Second Survey	Third Survey
Primary	3.5	8.2	6.5
Junior High	25.2	28.6	16.1
Senior High	30.1	38.8	41.9
Degree	34.5	22.5	32.3
Postgraduate	0.9	2	3.2
Total	100	100	100

These data also show that in each of the three surveys there were a higher percentage of pregnant women who believed that the academic achievement of their children would reach the high school level than the percentage of those who believed they would only achieve a basic level of education: at baseline it was nearly 2% (30.1% -28.7%) higher than in the second survey at 2% (38.8% -36.8%), yet in the third it was almost 20% (41.9% -22.6%). Meanwhile, the proportion of pregnant women that said that their children would achieve some level of college or postgraduate education was 35.4% at baseline, 24.5% in the second stage of data collection and 35.5% in the third (Table 39).

5.6 Fathers

The fathers of the children were also considered important to the design and implementation of the intervention. As caregivers, fathers attended early education sessions where they receive information on parenting practices to help improve their skills in child development, child care and protection and personal and community development. Aligned to this design, the questionnaire asked fathers about their parenting practices, attitudes inside and outside the home.

It should be noted that as in the questionnaire designed for pregnant women, it was a requirement that for the male parent to be interviewed in the third survey he had to have answered one of the earlier surveys, which is why the number of observations is reduced significantly. In this sense, we see that of respondents to the third survey, only 25% had responded in the baseline, and 40% in the second survey and 35% responded in both visits. During the first survey, parents had an average of 23 years of residence in the town, while the second this increased to 27, however the average remained constant for the third.

Among others, the effective time fathers spend with the family is an important practice that was included in the instrument and was measured by how often they ate with the child and the mother. Here we found that 84% did so once a day or more in the baseline, however, this proportion increased to 87% in the second survey and 92% in the third. Fathers who ate with the family few times per week represented 10% of the total in the first visit. This was reduced to 5.4% for the second survey, but reached 8% for the third. Finally, at baseline just 5% of fathers reported eating with the family once a week or less, in the second survey this percentage increased to 7.2%, while in the third stage, neither parent mentioned eating with their children with such frequency (Table 40).

Table 40: How often does [….] eat together with you and his mother?

	Percentage		
	Baseline	Second Survey	Third Survey
More than once per day	53.79	67.27	73.3
Once per day	30.3	20	18.7
A few times per week	10.61	5.45	8
Approximately once per week	3.79	5.45	0
Approximately once or twice per month	1.52	1.82	0
Total	100	100	100

Part of the instructions regarding parenting practices regarded preventive activities, entertainment and home support. In the first category we see the trend of parents who take their children to preventive medical check ups was downward throughout the study: baseline 63.2% of parents said that they did so, 62.5% in the second survey and 51.3% in the third. With entertainment parents registered their greatest participation in all three stages of the study. Almost all of the parents said they played with their children in the three surveys. The percentages of parents who said they sing with their children were lower than those who play, but in no survey the percentage was below 82%. Meanwhile, 97% of parents said they carried their children during baseline, 92.9% in the second survey and 89.3% in the third (Table 41).

Table 41: I am now going to ask you to please tell me if the following activities or situations occur:

	Percentage		
	Baseline	Second Survey	Third Survey
Take [….] for medical check ups when not sick	63.6	62.5	51.3
Plays with [….]	99.2	98.2	98.6
Sings with [….]	82.4	94.6	85.3
Helps prepare meals for [….]	72	78.6	73.3
Helps feed [….]	89.4	92.9	88
Helps dress [….]	81.8	87.5	82.7
Carries [….]	97	92.9	89.3

Regarding the category of home support practices, we see that the proportion of parents who helped prepare meals for the children remained similar throughout the study: 72%, 78.6% and 73.3% for the baseline, second and third stages, respectively. A supplement in this respect is given by the father's participation in the practice of feeding his child, an activity in which 89.4% of parents reported doing in the first survey, 92.9% in the second and 88% in the third. Another factor in this category was the support in the task of dressing their children in which 81.8% of parents reported doing so in the first survey, a percentage slightly higher in the second (87.5%) and (82.7%) in the third stages (Table 41).

Table 42: Health risk behaviors, socially and within the home

	Percentage		
	Baseline	Second Survey	Third Survey
Smokes	9.6	16.1	18.7
Decides on what to spend the household Money without asking	37.4	42.7	33.3
Has financial problems	46.2	55.4	61.3
Speaks with family to solve problems	94.7	94.6	97.3
Speaks with someone when sad	83.3	85.7	85.3
Speaks with someone when angry	54.4	58.9	58.7

Research was also done regarding health, family and social risk behaviors. In the first survey you can see that the percentage of parents who reported smoking increased at each stage of the study: 9.6% at baseline, 16.1% in the second survey and 18.7% in the third. With respect to household behavior, the report shows that the degree of communication in spending practices increased through the stages, as parents who said they did not participate in what money was spent on went from 37.4% at baseline to 33% at the end. The trend of parents who reported talking to their families was similar; in the first two stages 94% was reached, while reaching 97.3% in the third. As far as social behavior is concerned, we found that the proportion of parents who mentioned talking to other people when they feel sad ranged between 83% and 85% throughout the study, whereas those who reported talking when they felt angry was between 54.4% and 58.9% (Table 42).

Table 43: Up to which level of education do you believe [….] will study?

	Percentage		
	Baseline	Second Survey	Third Survey
Primary	3.1	3.6	0
Junior High	22.1	16.1	17.3
Senior High	19.1	48.2	28
Degree	51.9	32.1	52
Postgraduate	3.8	0	2.7
Total	100	100	100

Finally, an interesting finding among the three surveys was that expectations of educational attainment of both caregivers and parents, showed an upward trend in the study period. During the baseline survey, 25% of parents expected the child to reach some level of basic education, a figure that dropped to 19% in the second survey and 17% in the third. In regard to senior high school, we see that almost 20% of parents in the first year expected their child to reach this level while in the second this increased to 48% and the third the expectation was reduced to 28%. The biggest change occurred in higher education as 55% of the participants in the first visit expected the child obtain an undergraduate or graduate degree, a proportion that dropped to just 32% in the second yet again reached 55% in the third (Table 43).

Part II. Impact Analysis of Primary Caregivers and Children and Balancing of Groups

6. Take Up

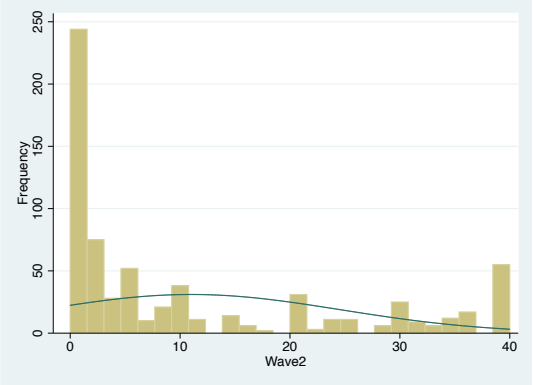
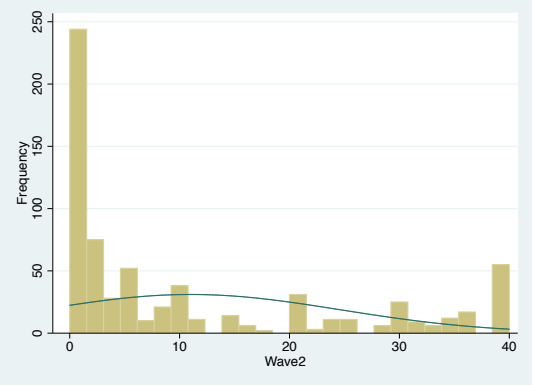
In order to understand the impact of the program on outcomes, it is first crucial to understand the impact on outputs. Based on interviews with caregivers, we calculate attendance at the Early Childhood program meetings. In a given year, promoters could facilitate up to 40 meetings per group. However, an examination of the data suggests that actual attendance was significantly lower (Table 44 below). At baseline, on average, treatment households had attended two meetings, which is consistent with the fact that there were initial organizational meetings in both treatment and comparison communities. At the end of the first year, caregivers reported attending – on average – 11 meetings. At the end of the second year, they reported attending just under 9 meetings on average.

Table 44: Intermediate results: take-up

How many meetings have you attended since we last spoke? (Treatment only)				
	Mean	Median	25th percentile	75th percentile
Baseline	2.1	1	0	3
End of Year 1	11.0	4	0	20
End of Year 2	8.7	3	0	12

However, closer examination of the data reveals an interesting finding: There is significant concentration of responses at 20 meetings and at 40 meetings per year (Table 45 below). This may reflect participants simply saying to enumerators that they attend every week (i.e., 40 meetings) or “about half the time” (i.e., 20 meetings). These high numbers pull up the average, when the median number of meetings is much more modest (Table 44). The median household attended only 4 meetings during the first year and 3 during the second year. Given this, any subsequent findings on child development or caregiver practices will be particularly striking.

Table 45: Caregiver Attendance at meetings, first year and second year

Attendance Year 1	Attendance Year 2
	

7. Impact on Caregiver Practices

One of the two principal goals of the program was to influence caregiver practices. To estimate the impact of this, there were a set of practices that the enumerator observes during the course of an interview (i.e., they are not reported by the caregiver). Different practices are queried for children age 0-3 and for children age 3-6. (The cut-off between the two groups is when a child turns 3.) We also use principal components to generate an index of all 8 observed practices observed (for children age 0-3) and all 12 observed practices (for children age 3-6).

We report four randomly selected practices for each age group as well as the index of observed practices (Tables 46 and 47). We observe a significant positive impact for one individual caregiver practice in each wave among younger children (Table 46), as well as in the overall index at the end of the first year (see the final column). At the end of the second year, the magnitude of the observed practice index is still large, but the result is very imprecisely measured. For caregivers of older children, the only significant impact is still “positive” (i.e., caregivers are less likely to restrict a child’s actions), but there are no significant impacts in the second follow-up, nor in the observed practice index.

Table 46: Impact on Observed Parenting Practices – 8 practices + an index – age (0-3)

	Spoke to child	Responded to child	Hugged or kissed child	Interfered child's actions	Observed practice index
Treatment 1st follow-up	0.05 (0.05)	0.14** (0.06)	0.06 (0.06)	-0.03 (0.05)	0.48** (0.20)
Treatment 2nd follow-up	0.01 (0.07)	-0.04 (0.07)	0.12* (0.07)	-0.01 (0.05)	0.28 (0.26)
Intercept	0.72	0.71	0.70	0.17	-0.02

Table 47: Impact on Observed Parenting Practices – 12 practices + an index – age (3-6)

	Spoke to child 2x	Responded to child	Hugged or kissed child	Restricted child's actions	Observed practice index
Treatment 1st follow-up	-0.03 (0.09)	-0.05 (0.09)	-0.03 (0.11)	-0.06* (0.04)	-0.18 (0.37)
Treatment 2nd follow-up	-0.12 (0.09)	-0.08 (0.09)	0.01 (0.11)	-0.04 (0.03)	-0.45 (0.38)
Intercept	0.75	0.80	0.58	0.03	-0.18

In addition to direct observation, we asked caregivers to report activities with the children, as seen in Table 48 (for caregivers of the youngest children) and Table 49 (for caregivers of older children) below. For the sake of brevity, we report the magnitude and significance level only. For caregivers of the youngest children, we observe an increase in children's books at the first follow-up albeit not at the second follow up. Note that the coefficient reports the difference in children's books between the treatment and comparison groups, not the actual level of books. At baseline, children age 0-3 had 1.5 books, and children age 3-6 had 1.8 books.

We also observe a decrease in reported television viewing. The television variables are normalized, so the coefficients suggest a substantial decrease of 0.23 standard deviations on weekdays and 0.41 standard deviations on weekends. As context, caregivers reported 0.8 hours of television on weekdays and 0.6 on weekends. For caregivers of older children (Table 49), the only significant difference observed is in the number of spankings, which is reduced in the treatment group at the first follow-up.

Table 48: Impact on Reported Parenting Practices – Brief Summary (sign/significance)- Age (0-3)

	1st follow-up	2nd follow-up
How many children's books?	1.17***	0.00
How often do you read or tell stories?	-0.15	-0.10
How often do you take […] to market?	0.11	-0.03
How many dolls or stuffed animals?	-0.03	-0.06
How many toys to push or pull?	-0.08	-0.03
How much TV watched on weekdays?	-0.23**	0.11
How much TV watched on weekends?	-0.41***	-0.10
How much is TV on?	-0.06	0.30
How often does […] eat with a parent?	-0.14	-0.18
How many spankings in the last week?	0.09	-0.21

Table 49: Impact on Reported Parenting Practices – Brief Summary (sign/significance)- Age (3-6)

	1st follow-up	2nd follow-up
How many children's books?	0.06	-0.21
How often do you read or tell stories?	-0.21	-0.25
How often do you take […] to market?	-0.36	-0.42
How many magazines at home ?	-0.09	-0.05
Do children listen to tapes or discs?	0.14	0.08
How much TV watched on weekdays?	0.05	0.09
How much TV watched on weekends?	0.10	0.05
How much is TV on?	-0.39	-0.32
How often does […] eat with a parent?	-0.11	0.14
How many spankings in the last week?	-0.66*	-0.16

This pattern of results suggests that the program does affect caregiver practices, despite the fact that caregivers attended the program very little. However, there is significant evidence of fade-out: Results are more significant and more consistent across measures

at the end of the first year. It is possible that with improved implementation and increased take up, that these effects would be more lasting.

8. Impact on Child Development

The second principal goal of the program is to affect child development directly. Here we report the impact on each of the five measured areas of the Ages and Stages child development instrument: Communication, gross motor skills, final motor skills, problem solving, and social skills. We then sum up the points for the five areas and create a standard normal index of overall child development. The results are reported in Table 50. After the first year, we observe (marginally) significant impacts in communication and gross motor skills, but not in the other areas. The impact on the overall index is positive but not statistically significant. At the end of the second year, results on communication and gross motor skills are still positive but are smaller and are no longer statistically significant.

Strikingly, in focus groups with promoters carried out both before the onset of the evaluation and later, the promoters reported that the child development areas in which they most expected to see improvements as a consequence of the program were communication and gross motor skills. While the program is intended to affect all of these areas, principal and initial activities are mostly focused around these first two.

Table 50: Impact on Child Development

	Communication	Gross motor	Fine motor	Problem solving	Social	Total
Treatment 1st follow-up	0.15* (0.09)	0.15* (0.08)	0.08 (0.09)	0.01 (0.09)	0.04 (0.08)	0.11 (0.09)
Treatment 2nd follow-up	0.09 (0.09)	0.08 (0.08)	-0.06 (0.09)	-0.08 (0.09)	0.00 (0.09)	0.01 (0.09)
Intercept	0.00 (0.02)	0.00 (0.02)	-0.01 (0.02)	-0.01 (0.02)	0.00 (0.02)	-0.01 (0.02)

9. Robustness and Heterogeneous Treatment Effects

All of the above analysis is carried out with all children who appear in the baseline, including both those who appear in Baseline and at the end of Year 2 as well as those that appear in Baseline and only at the end of Year 3. We replicated the analysis including only that smaller sample of children who appear in all three waves of data, and the pattern of results is the same.

When this research was initiated, we proposed to examine heterogeneous treatment effects based on the gender and age of the child, among other areas. Analysis based on the gender of the child reveals that the size of the impact was much larger for girls than for boys. Table 51 reports the treatment effect for each gender and the significance. For communication, girls improved more than 0.2 standard deviations at the end of the first year, and while the effect shrank somewhat by the end of the second year, is remained statistically significant. Likewise, the effect on gross motor skills is driven by girls. In fact, the overall index increases by 0.2 standard deviations for girls at the end of the first year. There is a surprising, inexplicable negative result on fine motor skills for boys by the end of the second year.

Table 51: Heterogeneous Effects on Child Development Based on Gender of Child

	Communication	Gross Motor	Fine Motor	Problem Solving	Social	Total
Wave 2						
Boy	0.09	0.12	0.03	-0.03	-0.07	0.04
	(0.10)	(0.09)	(0.10)	(0.09)	(0.09)	(0.10)
Girl	0.22**	0.18*	0.13	0.06	0.15	0.20*
	(0.10)	(0.09)	(0.10)	(0.10)	(0.09)	(0.10)
Difference (P-value)	0.07	0.38	0.28	0.17	0.00	0.02
Wave 3						
Boy	0.03	0.07	-0.21**	-0.11	-0.04	-0.08
	(0.10)	(0.08)	(0.10)	(0.09)	(0.10)	(0.10)
Girl3	0.17*	0.08	0.11	-0.05	0.04	0.10
	(0.10)	(0.08)	(0.10)	(0.10)	(0.09)	(0.10)
Difference (P-value)	0.06	0.94	0.00	0.39	0.28	0.01

Heterogeneous analysis by age suggests differential effects as well. Specifically, the results seem to be driven by older children (Table 52). The cut-off is 22 months at baseline, which was the median age of children at that point. Both communication and gross motor skills seem to have an effect driven by older children, and the gross motor skill effect endures in the second wave. As with the analysis by gender, there is a surprising, negative result on fine motor skills for younger children.

Table 52: Heterogeneous Effects on Child Development Based on Baseline Age of Child

	Communication	Gross Motor	Fine Motor	Problem Solving	Social	Total
Wave 2						
Older	0.43***	0.27***	0.05	-0.01	0.12	0.24**
	(0.10)	(0.09)	(0.10)	(0.10)	(0.09)	(0.10)
Younger	-0.15	0.01	0.10	0.03	-0.06	-0.02
	(0.10)	(0.09)	(0.10)	(0.10)	(0.09)	(0.10)
Difference (P-value)	0.00	0.00	0.61	0.57	0.02	0.00
Wave 3						
Older	0.14	0.22***	0.10	-0.20	0.07	0.10
	(0.10)	(0.08)	(0.12)	(0.10)	(0.06)	(0.10)
Younger	0.05	-0.06	-0.21**	0.03	-0.09	-0.08
	(0.10)	(0.09)	(0.09)	(0.09)	(0.10)	(0.10)
Difference (P-value)	0.25	0.00	0.01	0.01	0.01	0.02

Future analysis will include age and gender within the same regression to capture multicollinearity, as well as analysis of sample size within cells to discern whether surprising results may be driven by small samples.

10. Balance between the Treatment and Control Groups

With the aim of analyzing the data from a comparative perspective, we present the features of the treatment and control households. We also present childrearing practices of parents and caregivers, and the children's level of development over the three age ranges with the greatest frequencies. The main objective of this section is to demonstrate that a balanced distribution was maintained in the context and exogenous factors that affect the program. First, the differences that were statistically significant as reported in the first two surveys will be compared with the results of the third survey. Subsequently, the results of the latter will be presented.

Worthy of mention is that the differences presented below were calculated by simple arithmetic and the significance was obtained through lineal regression by which the variable depends on the treatment or control group. By nature, the variables below were not difficult to obtain or interpret the difference among the groups. Nonetheless, in cases where the variables are categorical, such as possession of goods, depending on quantity of measures, a value of 1 was assigned when the response was affirmative and 2 when negative. Consequently, the average represents a percentage of negative responses. With this type of transformation the percentages behave as a continuous variable, reason by which the interpretation of the differences can be estimated in a regression model and their behavior observed.

10.1 Household Balance

In the baseline report, on average the group of treatment households was higher than the comparison group. For the follow-up questionnaire, however, the difference between groups was not significant. The number of children qualified to participate in early childhood education program by age sustained a significant difference in both cases: during the baseline, household in the treatment group had a slightly higher average, but for the second survey this difference was of significance only at a confidence interval of 90% (Table 53).

In the first report, a significant difference in the flooring material for housing conditions of the treatment group was found but only for the baseline survey and not in the subsequent two. However, the difference persists in the management of waste in households: nearly 35% of households in the comparison group were connected to the public sewer system,

versus 24% of the treatment group; this is a significant difference even with a 1% margin of error (Table 53).

In terms of appliance ownership, baseline results showed that in the homes of the treatment group, the proportion of coal stoves and electrical power meters was higher than that of the comparison households. The first follow up survey results of comparison households reported a greater ownership of wood or coal stoves, a significant result with 95% confidence interval; while the difference in the possession of electrical power meters was similar to the baseline survey, and significant with a 99% confidence interval factor. In the third survey, the difference between groups of both variables decreased and lost significance (Table 53).

In the case of the second survey, the comparison group had a higher proportion in the possession of five specific appliances: water heater, shower and computer were significant at a confidence level of 99%; in the case of households that reported having Internet service, the difference is significant at 95%, and for those with a radio, the difference is only relevant with 90% confidence interval. For the third lifting said differences remained significant, except for the possession of computer. It was significant with a confidence interval of 99% for possession of shower; 95% for radio and internet, and 90% for the water heater. In addition, 61% of control households reported having a cell phone, versus 55% in the treatment group; this difference was significant with 95% confidence interval (Table 53)

Table 53: Differences Between Control and Treatment Households According to Panel Survey's Waves

	Differences between Control and Treatment		
	Baseline	**Second Survey**	**Third Survey**
Number of members	0.7178***	0.094	0.12***
Under 60 months	0.0336***	0.091*	-0.01
Pregnant Women	0.0049	0.026	-0.007
Members that speak Spanish	0.0002	0.000	0.014*
Members that can read and write	0.0001	0.001	0.003
Housing conditions			
Cement roof	1.1113	0.029	-0.033
Dirt floor	0.0452*	0.010	0.004
Walls	0.1026	0.003	-0.006***
Total rooms	0.0431	0.183	-0.134
Total bedrooms	0.1551***	0.100	-0.019
Days with electrical power	0.0094	0.060	-0.127
Drainage	0.2052**	0.104***	-0.137***
Goods			
Gas stove	0.0345	0.055	-0.048*
Wood stove	0.0556*	0.069**	0.029
Water tank	0.0148	0.016	0.011
Boiler	0.0170	0.089***	-0.052**
Cistern	0.0170	0.016	0.026
Shower	0.0355	0.115***	-0.108***
Power meter	0.0434*	0.115***	-0.022
Automobile	0.0332	0.018	-0.01
Internet	0.0085	0.024**	-0.023**
Computer	0.0170	0.039***	-0.02
Telephone	0.0138	0.018	-0.01
Cellular	0.0167	0.043	-0.06**
Washer	0.0061	0.049	-0.009
Refrigerator	0.0007	0.034	-0.043
Television	0.0090	0.021	-0.013
Radio	0.0069	0.061*	-0.067**

10.2 Caregiver Balance

A difference found in the baseline report was that treatment households had a higher rate of Oportunidades recipients; this difference was not sustained with first tracking data, but again became relevant in the third, with 58 percent of treatment caregivers reporting that someone in the household was a recipient, compared to 51 percent in the control group. In the case of caregivers with children aged 36-42 months, the comparison group had both parents living together more often; this difference was reversed in favor of the treatment group during the second stage, and with a confidence interval of 90 percent but lost significance on the third visit (Table 54).

On the first visit we found that comparison group caregivers reported bringing their children to historical or cultural sites more frequently; on the second visit this difference was not significant and the last visit this proved to be significantly in favor of the treatment group with a confidence interval of 99 percent. In the same age range, the comparison group caregivers allowed children to choose what they wanted to eat more frequently in the first year, but that difference was not maintained in the next two years (Table 54).

In addition to these changes, new differences emerged on the second visit: the caregivers of the treatment group reported that someone helped them care for the child more often; this result is significant with a confidence interval of 95 percent. This was reduced to 90 percent of confidence on the third visit. It was also more common for the treatment group interviewed to affirm that the home receives financial support from people living within the country, although this difference only holds a 90% confidence interval; in the third survey the difference became insignificant (Table 54).

Among caregivers with children 36 to 60 months, the comparison group reported more often that the child had a device for listening to music, although this result only holds for a confidence interval of 90 percent. However this lost significance for the third survey. It was also more frequent for comparison group caregivers to talk to the child when they were busy, holding the same level of confidence interval. However significance was again lost in the last visit. In addition, comparison group caregivers reported more frequently that the child feels somewhat or very close to his father, a significant difference with a 95% level of confidence interval; again, this difference was lost for the third survey (Table 54).

In the same age group, only two differences in favor of the treatment group were presented during the second survey. Firstly, it was more common for children to eat with their father once a day or more, and this difference is supported by 99 percent confidence interval. Also, the average spanking the caregiver had given a child in the week prior to the survey was lower, and the difference is significant for a confidence interval of 95 percent. However, both differences were not statistically significant for the last survey. In Table 54 the total results for the survey of caregivers by study group is presented.

Finally, there were four differences that had emerged in the third survey compared with the prior visits. First, the comparison group caregivers reported being recipients of Seguro Popular more frequently, though the difference is only significant by a confidence interval of 90 percent. Second, among caregivers with children older than 36 months, the treatment group helped their children learn numbers more often; this difference was significant for a confidence interval of 95 percent. Third, 84 percent of caregivers in the treatment group reported that the father sang with his child, versus 78 percent in the comparison group; the probability of error in this difference was less than 1 percent. Fourth, comparison group caregivers planned to enroll their children in preschool at an average of 3.7 years of age, while the treatment group average increased to nearly 3.8 years of age; however, this difference was significant with a confidence interval of 90 percent only (Table 54).

Table 54: Differences Between Caregivers in Control and Treatment Households According to Panel Survey's Waves

	Differences between Control and Treatment		
	Baseline	**Second Survey**	**Third Survey**
Age of child	0.0233	0.019	0.078
Time living in the community	0.9353	0.853	-0.508
Mentioned a second caregiver	0.0304	0.052**	-0.047*
Oportunidades recipient	0.0763***	0.061	0.065**
Seguro Popular recipient	0.0121	0.017	0.048*
Procampo recipient	0.0095	0.027	0
Scholarship recipient	0.0120	0.031	-0.026
Senior citizens recipient	0.0063	0.003	0
Economic support from someone in the country	0.0033	0.037*	0.029**
Economic support from someone outside the country	0.0016	0.016	-0.012
Was treated in a government clinic	0.0304	0.003	0.026
0 to 36 months			
Children's books	0.0610	0.38***	0.022
Stories	0.4009***	0.042	0.08
Plus toys	0.0252	0.002	0.028
Pull toys	0.0803	0.025	0.06
Parents live together	0.0066	0.013	0.008
Frequency with which child eats with father	0.0849	0.018	0.038
Frequency with which speaks with child	0.0936	0.01	0.064
Closeness of child with father	0.0241	0.007	0.092
Closeness of child with mother	0.0176	0.019	-0.018
Corporal punishment	0.0808	0.074	-0.02
36 to 42 months			
Children's books	0.0944	0.025	-0.034
Stories	0.6328***	0.044	0.045
Magazines	0.0773	0.041	-0.043
Device for listening to music	0.0255	0.077*	-0.039
Helps child learn numbers	0.0091	0.008	0.032**
Helps child learn letters	0.0170	0.037	0.022
Helps child learn colors	0.0432	0.028	0.015
Helps child learn shapes and sizes	0.0584	0.024	0.002
Child chooses what to eat	0.3534**	0.005	-0.012
Visits museums and historical sites	0.3011**	0.023	0.033***

	Differences between Control and Treatment		
	Baseline	**Second Survey**	**Third Survey**
Parents live together	0.1105**	0.049*	-0.003
Frequency with which child eats with father	0.1921	0.098***	0.024
Frequency with which speaks with child	0.0277	0.078*	0.016
Closeness of child with father	0.0501	0.076**	0.014
Closeness of child with mother	0.1456**	0.017	-0.016
Corporal punishment	0.0171	0.524**	-0.086
Chile rearing practices			
Medical check ups	0.0122	0.009	0.031
Plays with child	0.0150	0.022	0.016
Sings with child	0.0277	0.021	0.063***
Musical instrument	0.0058	0.017	-0.018
Child illness	0.0419	0.033	0.019
Educational achievement expectations	0.0665	0.001	-0.064
Preschool registration	0.0065	0.003	0.002
Age at preschool registration	0.0122	0.186	0.085*

10.3 Pregnant Women Balance

The only significant difference reported in the baseline questionnaire was in the length of time residing in the town. However, in this questionnaire the person that answered the questionnaire could not be another person than the one who answered the previous questionnaire; this question was eliminated for the follow-up questionnaire. For the second survey, two new differences between the groups were identified: first, mothers in the comparison group mentioned someone who helped them in the care their child most often, a difference that is significant to a confidence interval of 95%. Furthermore, in the same group of mothers it was more common to exchange tips with other pregnant women, although this result was only maintained with 90% confidence interval. The rest of the questionnaire showed no significant differences in variables (Table 55).

Table 55: Differences Between Pregnant Women in Control and Treatment Households According to Panel Survey's Waves

	Differences between Control and Treatment		
	Baseline	**Second Survey**	**Third Survey**
Economic Support			
Oportunidades recipient	0.019	0.193	-0.086
Seguro Popular recipient		0.241	-0.086
Procampo recipient		0.108	-0.091
Scholarship recipient		0.108	-0.064
Support from persons living in another country		0.037	-0.064
Months since delivery		0.952	0.015
Care during pregnancy			
Medical check ups	0.160	1.568	0.002
Non prescribed medicine	0.037	0.013	-0.018
Alcoholic beverages	0.051	0.04	0.015
Coffee	0.087	0.087	0.069
Soft drinks	0.269	0.062	0.051
Someone smokes inside the dwelling	0.021	0.042	-0.049
Change of diet	0.115	0.176	0.167
Care after pregnancy			
Breast feeding	0.004	0.043	-0.063
Months of lactation	0.974	1.708	-0.1
Babies are aware of what is going on around them	N/A	0.008	-0.087
Preschool registration	1.022	0.111	-0.032
Expectation to achieve more than a Basic education	0.174	0.15	0.06

10.4 Balance of Fathers

In the baseline questionnaire, two differences between the groups were presented: the treatment group fathers most frequently stated that they helped to feed the child; also the same group responded more often that the child felt somewhat or very close to them. However, neither difference is supported by follow up data.

Besides these, data from the second survey yielded two new differences: 25% of the comparison group fathers said they had argued with their partner in the last week, while no one in the treatment group reported arguing; this difference is significant for a confidence interval of 95%. Also, all treatment group fathers said they talk about their children with other parents, compared with 92% of the comparison group, and this difference holds for a confidence interval of 99% (Table 56).

Table 56: Differences Between Fathers in Control and Treatment Households According to Panel Survey's Waves

	Differences between Control and Treatment		
	Baseline	**Second Survey**	**Third Survey**
Time living in the community	1.366	2.1375	-1.498
Medical check ups	0.015	0.175	0.222*
Plays with child	0	0.025	0.025
Sings with child	0.026	0.075	-0.055
Helps in preparing child's meals	0.119	0.05	0.057
Helps to feed the child	0.169***	0.0125	0.112
Helps to dress the child	0.096	0.0875	0.102
Carries child	0.024	0.1	0.033
Closeness to child	0.376***	0.075	0.063
Educational achievement expectations	0.156	0.1	0.059
Preschool registration	0	0.05	0.067
Social behavior			
Smokes	0.067	0.05	-0.019
Smokes in the home	0.286	0.142857	0.041
Decides spending without asking	0.020	0.075	-0.018
Has financial problems	0.151	0.1875	0.062
Speaks with family to solve problems	0.061	0.075	-0.005
Speaks with family when sad	0.033	0.025	0.161*
Speaks with family when angry	0.083	0.05	-0.051
Alcoholic beverages	0.002	0.0125	-0.023
Argues with spouse	0.036	0.25**	-0.019
Speaks about children with other parents	0.223	0.075***	0.203*

10.5 ASQ Balance

Unlike the previous sections, the results of the Ages and Stages Questionnaire 3 cannot be aggregated into a single measure of the state of development of each child evaluated, but provides the score obtained in each section independently, this is due to different cut off lines are for each age range and area of evaluation. Given the complexity of presenting the differences of the five areas for each of the twenty questionnaires used, only three are presented. In the previous two reports, the ASQ results for children aged 2, 6 and 12 months were used as examples; however, due to the growth of children assessed as in this survey, no new children were included. The youngest group in this report is 14 months, and even this group provides few observations. In order to show the most relevant results, we have selected age groups with more than 100 children, being the groups of 30, 42 and 54 months.

In the first age range, we found the treatment group presented an average of one percentage point higher in communication skills, while the comparison group showed better development in gross motor skills, fine motor skills and problem solving abilities. However, none of these differences were statistically significant (Table 57).

Table 57: ASQ results 30 months per group for the third survey

	Participation	Comparison	Difference
Communication	45.686	44.729	0.957
Gross Motor	50.098	51.756	-1.658
Fine Motor	40.588	43.243	-2.655
Problem solving	34.411	37.162	-2.751
Social	47.352	47.162	0.19

For the age range of 42 months, the treatment group scored highest than the control group in gross motor and social areas; the latter, significant to a confidence interval of 95 percent. The comparison group, however, had better results in communication, fine motor and problem solving skills, but not statistically significant (Table 58).

Table 58: ASQ results 42 months per group for the third survey

	Participation	Comparison	Difference	
Communication	45.796	46.818	-1.022	
Gross Motor	50.973	50.858	0.115	
Fine Motor	37.212	38.535	-1.323	
Problem solving	41.15	42.626	-1.476	
Social	50.796	47.727	3.069	**

In the 54 months range, the treatment group performed better in the assessment of communication, gross motor and fine motor skills. Moreover, the comparison group had higher scores on problem solving and social skills; however, no results were significant (Table 59).

Table 59: ASQ results 54 months per group for the third survey

	Comparison	Participation	Difference
Communication	50.921	50.103	0.818
Gross Motor	55.964	55.463	0.501
Fine Motor	46.667	46.082	0.585
Problem solving	35.131	35.876	-0.745
Social	48.815	50.927	-2.112

Conclusions

The main objective of the impact assessment study of the Early Childhood Education Program was to generate reliable and valuable information for decision-makers, mainly related to education policy. Based on this objective, a three-stage longitudinal study whose cornerstone was the collection of quantitative data was designed that relied on information gathering instruments to compile data on caregiver child-rearing practices and the cognitive development of their children.

With the target of providing value and reliability to the study, the surveys were structured to reduce the bias of the data as best as possible. In this sense, a system was implemented to improve the information gathering processes, which ensure adherence to the code of ethics and data collection protocol developed by CIDE's researchers. Special importance was placed the processes of adjusting questionnaires, training field researchers, survey supervision, coordination with authorities of CONAFE and contractors, as well as the capture and validation of database contents.

Five factors of vital interest to the study were jointly observed. One of these was the balance between the control and treatment groups for the longitudinal design, which provided confidence interval in estimates of the differences between the groups regarding the variables of interest as well as the final impact. The second factor was that during this latest survey, the results observed were relatively consistent with those obtained during the baseline and the second wave of the panel. The third demonstrates that the relative response rate, especially for treatment communities increased for the third stage, which provided greater confidence interval for impact estimates due to sequential care during the every process of the evaluation.

Finally, the most important factor: after having performed the activities planned, we are pleased to have observed positive and significant differences in both the practices of primary caregivers, and the cognitive development of their children in the treatment groups versus the control groups. Together, these factors point to two issues of great importance to be considered in research projects such as this: data confidence interval is vital and with it, the assurance that these impact estimates provide greater accuracy, enabling decision makers of early childhood education policy.