A Randomized Impact Evaluation of Early Childhood Development in Rural Mozambique: follow-up 6 years after Data documentation

General background information

Starting in 2008, Save the Children implemented a center-based community driven preschool model in rural areas of the Gaza Province of Mozambique. The project financed the construction, equipment and training for 67 classrooms in 30 communities, to provide Early Childhood Development (ECD) activities for children aged between 36 and 59 months. . As part of its design, the program included an experimental impact evaluation (using Randomized Controlled Trial) whereby the 30 intervention communities were selected at random from a pool of 76 eligible sites.

Before the preschool activities initiated, a baseline survey was carried out in 76 communities in Gaza Province across the districts of Xai-xai, Bilene, and Manjacaze.

Two years later, in 2010, the same 2,000 households participated in a mid-line survey to evaluate the impact of the program after one or two years of potential exposure to pre-school.

The present data correspond to the follow-up survey that took place in 2014, namely 6-years after the beginning of the intervention when the targeted children were supposed to be in primary school age. The same data collection firm, COWI (called AustralCOWI before 2014), was in charge of the baseline (2008), midline (2010) and follow-up (2014).

The impact evaluation has four main research questions: (1) to evaluate the efficiency of low-cost community-based preschool program in a disadvantaged rural African setting in terms of cognitive, socio-emotional skills as well as learning outcomes for the children, (2) to evaluate the effects of such intervention on school enrollment, attendance, and progress (i.e. grade promotion, repetition, dropout); (3) to assess whether parenting practices and knowledge can be durably influenced by community-based ECD program; (4) To identify potential spill-over effects of the program on health, education, productivity and labor market outcomes of siblings and parents of preschoolers.

Field work was carried out from April to November 2014. In addition to household surveys and cognitive assessments of children, data from primary school directors, pre-school animators and community leaders were collected during this period. From the original 2,000 target children of the 2008 survey, more than 90% of them were successfully tracked and geo-referenced all over Gaza and Maputo City, and Maputo Province.

Questionnaires

In 2014, the following survey instruments were used: (i) a Socio-economic questionnaire, (ii) a children assessment in literacy, numeracy and non-verbal reasoning, (iii) a school questionnaire administered to directors in all schools of sampled communities, (iv) a community leader questionnaire, and (v) a questionnaires to collect information on the functioning pre-school centers in the treatment communities.

The learning assessment has been administrated in the household of the target child following the socio-economic questionnaire. 10 minutes were allocated to the Literacy part, 10 minutes to the Numeracy section and 5 minutes to the non-verbal reasoning section.

A more detailed description of the various instruments is presented below.

Instrument	Sections of the instruments
Socio-economic questionnaire	1. Household information
	2. Education Information on members between 5 and 18 years old.
	3. Job section (all members)
	4. Preschool history (for less than 12 years old members)
	5. Time-use section for the target child, on the activities of the day before the interview
	6. Contact section
	7. Condition of the interview.
Assessment of literacy, numeracy and non-verbal reasoning	*Literacy (administered in Portuguese only) :*
	Item 1 : Letters recognition (grading scale : /3)
	Item 2: Words recognition (grading scale /3)
	Item 3: Simple objects recognition (out of picture) (grading scale /4)
	Item 4: Reading out lout a simple sentence (grading scale :number of words read correctly /5)
	Item 5 : Reading out lout a text containing 59 words (grading scale :number of words read correctly /59)
	Item 6: Comprehension question on the text (in writing by the child) (grading scale /1)
	Item 7 : Comprehension question on the text (in writing by the child) (grading scale /1)
	Item 8: Comprehension question on the text (multiple choice among 3) (grading scale /1)
	Numeracy (administered in local language or Portuguese depending on how comfortable the child is) :
	Item 1: Number recognition (grading scale /3)
	Item 2 : Number ordering (grading scale /1)
	Item 3: Simple digit addition (grading scale /1)
	Item 4: Double digits addition (grading scale /1)
	Item 5: Triple digit addition (grading scale /1)
	Item 6: Simple digit subtraction (grading scale /1)
	Item 7: Double digit subtraction (grading scale /1)
	Item 8 : Simple digit multiplication (grading scale /1)
	Item 9 : Double digit multiplication (grading scale /1)
	Item 10: Triple digit multiplication (grading scale /1)
	Item 11: Simple digit (integer resulting) division (grading scale /1)
	Item12 : Division with 2-digits numerator and one digit denominator (with integer result) (grading scale /1)

	Item 13: Ordering fraction with common numerator (multiple choice) (grading scale /1)
	Item 14: Quick problem solving involving multiplication by ten. (grading scale /1)
	Item 15: Complete logical sequence(grading scale /1)
	Non-Verbal Reasoning :
	Item 1-4 : Non-verbal reasoning involving completion of logical matrices. (grading scale /1 for each items)
School Questionnaire	1. Director and school characteristics
	2. School enrollment data, by grade and gender
	3. School Infrastructure
	4. Interaction with parents and community
	5. Interaction between Primary School and ECD center (*Escolinha*)
	6. Subjective pupils' assessment
Community Leader Questionnaire	1. Information on the leader
	2. Information on existing Pre-school centers
	3. Estimated distances from basic facilities
	4. Information on local crops
	6. Social capital in the community
ECD Questionnaire	1. Instructor's Information
	2. Activities in Preschool
	3. Motivations and stipends
	4. Infrastructure

Sample:

Communities sampling-process (baseline)

The design used for this impact evaluation is that of a clustered randomized control trial (C-RCT) at community levels

Stage 1: Community Eligibility.
Within the three target districts, a subset of eligible communities is identified that meets two key operational requirements for implementation of the program:
1. Population size: To qualify for the intervention, communities must have a population no less than 500 and no more than 8000 people. This range was determined as operationally feasible given the community mobilization process that accompanies the establishment of each ECD center.
2. Clusters: Management of the intervention requires that the intervention be clustered in groups of 6 treatment communities that can be served by a program staff. The definition of cluster was set set by Save the Children, based on minimum criteria of operational feasibility (distance or time traveled between sites).

The complete universe had 252 villages in three intervention districts. After applying eligibility criteria of population size and clustering, the sample was reduced to 167 villages in 11 clusters.

Stage 2: Clusters selections
The largest clusters in each district were selected for inclusion in the sample, resulting in total of 98 villages. To achieve coverage in all three distracts, it was further agreed with the NGO that the sample would include 2 clusters each in Manjacaze and Xai Xai and one cluster in Bilene

Stage 3: Community level randomization
Within clusters of communities that meet the two requirements outlined in stage 1, communities form triplets based on population size, and from each triplet a treatment community is selected at random. The two smallest villages which did not form part of a triplet were dropped. The final sample is composed of 37 treatment (7 for replacement) and 59 control villages (11 replacement), for a total sample of 96 villages. A total of 30 new intervention communities were then selected for this round of implementation through random assignment.

No replacement of communities was needed.

Child level selection
In addition to randomization at the community level, there is exogenous variation in treatment within communities, based on rules of eligibility for Orphans and Vulnerable Children (OVC). ECD centers had a maximum of 3 class rooms with 35 students per class, for a maximum of 105 students per preschool. In the case of over-subscription of children to the ECD centers, Save the Children and the communities selected the children through a lottery system.

Household sample:
A total of 2,000 households with preschool age children were sampled from the 76 evaluation communities at baseline. With no household listing available at the time of the survey, a census of each community was carried out to identify households with children in the age range of 36 to 59 months. Taking the list of households with at least one child in this age range, 23 households per community were planned to be selected randomly. In addition, in 4 large treatment communities where oversubscription to the program was likely, an additional 63 households were selected, yielding a total sample of 2,000 households. In practice, some communities did not have 23 households eligible. In this case, all eligible households were sampled while in larger communities, more households than planned were sampled.

Endline tracking sample:

The tracking sample was carried out between April and November 2014, following 3 Phases.

Phase 1: Community Survey

a) Child and Household Survey: The household survey cover the panel of 2000 households with children that were aged 36 to 59 months at the time of the baseline survey in 2008 and who were followed in the first follow-up survey in 2010. This cohort was approximately 96 to 120 months at the time of this second follow-up survey. The survey also collected the Geo-Referenced coordinates of each target child's current place of residence using a GPS supplied by the firm to each surveyor.

b) School Survey: A survey on the panel of approximately 76 schools in the original study areas was conducted to collect school-level data on the performance (i.e. grades) and school progression (i.e. repetition and drop-out rates) of children in Grades 1 to 5.

c) Community questionnaire: In the 46 control communities, a brief community survey was administered to assess whether any preschool activities were implemented in this community since 2008. In the 30 treatment communities, additional questions were included to assess the extent to which the Save the Children program continued to exist after 2010.

Phase 2: Tracking of All Movers

d) Tracking Sample. At the end of the Phase I (June 2014) 1,607 households with targeted children were completed, and 383 target children had moved away, or have not been identified in their original locality of residence. With the objective of minimizing biases from sample attrition, all children who moved from their last known place of residence were tracked by the survey firm to their new locality of residence and surveyed. Tracking followed the target child, even if some of his or her relatives are found in the original location. Children that have moved within the original community were not considered movers, and were located and interviewed as part of the standard field work operation under activity Phase 1a. Only children that have moved to a new locality were eligible for inclusion in the household tracking sample. If the target child passed away between 2008 and 2014 the socio-economic questionnaire was still administrated (under the consent of the caregivers). At the end of the phase 2, 210 target children were successfully tracked.

Phase 3: Intensive Tracking Sample

In a final phase of tracking all the targeted children not located in phase 2 were intensively tracked with the objective of identifying the child's current location and completing the survey. Given budget constraint, area of tracking only included Maputo, Maputo province, Gaza province, and south of Inhambane province. Phase 3 tracked 1889 households. However 14 target children couldn't be matched with absolute certainty with the baseline, resulting in the use of 1,875 socio-economic questionnaires.

Reasons	Number
The whole family died	2
The target child died and the interview was not possible	5
The age of the target children is out of the study age range	5
The target child moved to Swaziland	1
Household moved to South Africa	15
Household moved inside Mozambique	15
Household moved without contact	24
Household not recognized in the community or target not identified	19
Household Refused	8

Repeated household	17
Total	**111**

Among the 1,875 successively tracked households, 1,830 target children were assessed with the literacy, Numeracy and Non-verbal reasoning test.

Field-work details:

The roll-out of the baseline and of the endline data-collection activities took place as follows:

Activity	Description
Socio-economic questionnaire	*Fieldwork :* Date: April 15 to September 17th, 2014 Number of households recovered: 1,889. Number of questionnaire administrated: 1,875 . *Data entry :* Date: May,1st to December 15th. 2014 Type: Manual entry using CS.Pro with double entry procedure Administrated to : the caregivers of the target child. In some specific case the caregiver was not a member of the household. In this case, questions in relation to the household were asked to the head of the household and questions in relation to the target child (education, time use, etc) were asked to the caregiver. The variable *caregiver_status* indentifies if the caregiver is a member of the household or not.

Assessment of literacy, numeracy and non-verbal reasoning	Date: April 15 to September 17th, 2014.
	
	Number of tests administrated in the household : 1,830
	Reasons for missing the test :
	    -> 30 children passed away
	    -> 2 refused to answer the test
	    -> 6 had disability and were not able to participate
	    -> 2 were sick at the time of the visit
	    -> 5 unspecified reasons
	
	Data entry (Idem)
	
	Administrated to : the target child only

School Questionnaire	Date: April 5 to September 12, 2014 Number of school visited : 56 Information on enrollment, repetition, and missing student were not evenly available in the schools.

	Administrated to : School directors and/or pedagogical directors of the any primary school located in the community.
Community-leader Questionnaire	Date: April 5 to September 12, 2014 Number of community leaders interviewed : 76 Due to misclassification in the national statistical census file in 2008 some neighbourhoods (bairros) were actually classified as communities. For those communities the community leader answered twice, referring to the specific parts of the community include in the sample. This happened in three communities (with 2 different neighbourhoods each). Administrated to : the person recognized as the leader of the community.
Pre-school survey	Date: April 5 to September 12, 2014 Number of preschool site visited: 28 Number of functioning preschool visited: Only 4 were still working and instructors were interviewed. When the preschool were not working, only pictures and GPS coordinated were taken. Two (non-functioning) preschool centers sites were not visited for unknown reasons. Administrated to : Preschool instructor if any.

Use of the data:

Socio-economic questionnaire:
Name of the questionnaire: SEQ_2014.dta

Treatment of missing data :
All variable from R05_2 to R22_15 recoded "." if missing.

The variable **Merge_panel** informs if the individual can be matched with the baseline(2008).

Data show limited inconsistency between the age reported in R07 and the age reported in EV2 for the target child taking the test. Reported variables on school history (R62, R25, R63) largely inconsistent with official age of entrance at primary school.

Reassignment of the treatment variable took place in 2008, because of the proximity of some control communities with treatment communities (see baseline and mideline reports).

Assessment of literacy, numeracy and non-verbal reasoning:
The results of the test are attached to the SEQ_2014.dta. One test was administrated to the target child for each household and results variables are reported for all the houshold members. The variable *target* identifies the target child (1 if target, 0 otherwise). The variable *valid_test* identifies if the test is considered valid for analysis (1 if valid, 0 otherwise).

Items were recoded "0" if the child did take the test but did not answer the exercise (originally coded "-4"). If the test, or some questions were not administrated variables are missing.

Items were graded "0" if incorrect answers.

School Questionnaire
Use the variable **COD_COMUNIDADE** to merge the dataset. Some schools can be associated with several communities.

Community-leader questionnaire
Use the variable **COD_COMUNIDADE** to merge the dataset.

Pre-school survey
Use the variable **COD_COMUNIDADE** to merge the dataset.

Weights
Two types of weights are included in SEQ_2014.dta. Both are at the community level and evaluate the inverse probability of selection within the community.

Variable "Weight1" has been generated using the number of eligible households (with at least one child aged between 36and 59 months) in the community (measured during the baseline fieldwork) divided by the number of households sampled in the community at baseline.

Variable "Weight2" has been generated using the size of the community as recorded in the national census in 2007 divided by the number of households sampled in the community at baseline.

Weights might or not be used for the analyses.

Constructed or additional data sets
There is one variable, among both the baseline and the endline data-sets, that was constructed by means of compounding measurements. The variable is called score_real, and it was built by summing up all the right answers to the Provinha test for each student (as each right answer is worth 1pt). It can be compared to the variable score_admin, which contains the results to the test as computed by the administrators of the test at baseline, and by IFP students at endline.

Treatment variables

Two different treatment variables are included in SEQ_2014.dta. The variable *treatment* identifies treated communities as originally planned. The variable *T1* identifies treated communities using ex-post reassignment given the proximity of some control communities with preschool centers.

Access
The conditions of access will be as follows.

- o   For the **internal staff-catalogue** (the Microdata library):
  - ✓ *Licensed* access through March 2018
  - ✓ *Public* access starting in March 2018
- o   For the **external public-catalogue**:

- For data collected before July 2015:
  - ✓ *No access* until the July of 2016
  - ✓ *Licensed* access through March 2018
  - ✓ *Public* access starting in March 2018

The name and contact information of the representative of the World Bank authorized to grant access to this information is: Marie-Helene Cloutier (mcloutier@worldbank.org, 202-473-5275); and Sophie Naudeau (snaudeau@worldbank.org, 202-473-7227)

Documentation:
List of documentation available :

- Socio-economic questionnaire
- School questionnaire
- Community leader questionnaire
- Preschool questionnaire
- Baseline reports
- Midline working paper

As the pupils assessment comes from a larger set of instruments from the Service Delivery Indicator survey, and given that it might be used in other context or countries the test item cannot be shared.

Codes
All codes are included in the questionnaires and/or labelling of the data sets.

Other

The time-use section of the socio-economic questionnaire is administrated to the caregivers and refers to the target child activities that took place on the day before the interview. See dictionary of variables for the use of this section.