The aeioTU Longitudinal Study
Project overview

Project Summary

The aeioTU Longitudinal Study is a randomized trial longitudinal study which enquires into the immediate and short-term cognitive, linguistic, nutritional and social effects of a comprehensive nutritional and educational 0-5 intervention in Colombia, in the form of aeioTU centers. The research is headed by the National Institute for Early Education Research (NIEER) at the University of Rutgers and local partner the Center for Economic and Development Studies (CEDE) at the Universidad de los Andes. The study is an independent evaluation carried out by academic researchers.

The project uses a two-site, randomized trial with a longitudinal design. Program quality is monitored. Analyses investigate main effects and interactions using a person-process-context model that includes family and community characteristics. Participants are 1217 low-income children and their families. The study addresses the following questions:

- What are the immediate and short-term effects of enrollment and participation in aeioTu on children's early learning and development?
- To what extent do these effects vary by intensity (cohorts)?
- Do pre-school effects persist over time, and in particular as children transition to primary?
- How do child and family characteristics moderate the effects on children?
- What is the cost-effectiveness of this intervention?

To date, NIEER has collected extensive background information on families, baseline child assessments, quality of care assessments, and observations on the home environment. Currently, NIEER is conducting a third follow-up of children's cognitive and non-cognitive development. We are presently seeking funding towards work to 5th year follow-ups that will allow the research team to evaluate the effects on school readiness and assess the transition to primary school for all children in the sample.

The study builds on i) findings from experimental studies in the U.S.; ii) recent evaluations in Latin America, the Caribbean, Africa, and Asia; iii) on-going early childhood studies in Colombia; and iv) socio-biological models of child development. The design allows for the estimation of the effects of dosage (years and quality of intervention). The aeioTU Longitudinal Study is the first of its kind in Latin America and unique for its longitudinal design which allows for measuring short and long term effects.

Design

The study is being conducted in two aeioTu centers in the northern coastal region of Colombia. We use a two-site, longitudinal, randomized trial. Sites vary in population density, poverty levels, incidence of violence, welfare services and health indicators. Eligible children are: underserved; poor children, qualified as SISBEN 1 and 2; at-risk due to nutritional status; victims of displacement due to conflict; or of indigenous origin. SISBEN is an index of economic disadvantage used for programs targeting in Colombia and categories 1 and 2 represent the two most vulnerable classifications out of 6. Households were randomly assigned from a waiting list and children assigned to aeioTu or the control group. Siblings receive the same assignment because households are the unit of randomization given our ecological framework.

The methodology embeds a cross cohort study to ensure that treatment and control groups are truly comparable at each age in measured family characteristics. Given that a smaller number of children is served at younger ages, we use stratified random sampling to construct samples in which there are no significant cohort (i.e. age-related) differences in observed family characteristics. Families were assigned to treatment, stratifying by age/gender groups and according to center capacity (n=320).

Random assignment strongly reduces the plausibility of most threats to internal validity (Shadish, Cook and Campbell, 2002), that is, any initial group differences in maturational rates, experiencing simultaneous historical events, etc. We administer the same assessments to all children regardless of treatment, under the same conditions (in spaces especially adapted for testing inside the communities). Baseline (or pretest) was carried out before randomization. Baseline data shows equivalence among treatment and controls in cognitive, social-emotional and nutritional dimensions, as well as in parental income, education, employment, welfare, family composition, immunization, assets and living conditions.

Instruments and indicators

Various instruments are used to assess different dimensions of child development. Linear models, hierarchical linear modeling, and classroom random effect models are used to estimate impact and the role of moderators. We use instruments that evaluate children's emotional, social and cognitive development over time. We also measure family characteristics, program costs (direct and indirect) and program quality.

Instruments vary with children's age, as cohorts progress through pre-k, kindergarten and into primary. We assess children's early cognition and motor skills, receptive vocabulary, emergent literacy, early math skills, and socio-behavioral abilities and schooling outcomes each academic year. We assess these specific abilities to observe the effects of treatment on brain development and cognitive capacities that emerge as various areas of the brain mature. Assessing specific abilities allows inferences concerning the periods of development and specific brain areas that are vulnerable, and may shed light on the biological and psychological mechanisms through which interventions affect child development.

Specific instruments and indicators used are:

i. **Family Background and Home Environment questionnaires**: Parents questionnaires collect information on parental ethnicity, parental income, savings, education, employment, welfare, family composition, immunization living conditions, parenting practices and information on the early childhood experiences.

ii. **Infant Development:** The Bayley Scales of Infant Development (BSID) are the most commonly used assessment of infant development (Fernald, Kariger, Engle & Raikes, 2009; Bayley, 2005). The Bayley has shown to predict later non-verbal and verbal cognition, i.e. as measured by the Test de Vocabulario en Imágenes Peabody (TVIP) (Blaga, Shaddy, Anderson, Kannass, Little & Colombo, 2009). As children grow, we measure child development using the Vineland Adaptive Behavior Scales (Sparrow, Balla, & Cicchetti, 1985). The Vineland is an individual parent questionnaire that assesses personal and social skills in communication, daily living skills, socialization, and motor skills.

iii. **Language:** The Test de Vocabulario en Imágenes Peabody (TVIP) (Dunn, Padilla, Lugo and Dunn, 1986) uses 125 items to assess receptive vocabulary of Spanish-speaking students. The TVIP has been used extensively in preschool studies (Early, Maxwell, Burchinal, Alva, Bender, Bryant, et al., 2007).

iv. **Math and Literacy:** The Woodcock-Muñoz III Tests of Achievement (WM-III) The WM-III is a comprehensive set of individually administered tests of children's early literacy and mathematical skills and knowledge, and we use subtests #1, #9 and #10, letter-word identification, text comprehension and applied problems, respectively (Muñoz-Sandoval, Woodcock, McGrew & Mather, 2005). In addition, the Early Literacy Skills Assessment (ELSA) (DeBruin-Parecki, 2005) measures four key elements of early literacy development–comprehension, phonological awareness, alphabetic principle, and concepts about print.

v. **Socio-emotional Development:** The Ages and Stages Questionnaires for the Socio-Emotional domain (ASQ: SE) (Squires, Bricker and Twombly, 2009) is a parent-completed assessment for children ages 6-60 months on socio-emotional development. The ASQ has been used for early development assessments in low and middle income countries (Handal, Lozoff, Breilh and Harlow, 2007; Tsai, McClelland, Pratt & Squires, 2006). As children grow older, we switch to the Behavior Assessment System for Children (BASC-II), which measures adaptive and problem behaviors (Bracken, Keith, & Walker, 1998; Doyle, Ostrander, Skare, Crosby & August, 1997).

vi. **Executive Function:** Head-Toes-Knees and Shoulders: The HTKS examines behavioral regulation (Ponitz, McClelland, Matthews & Morrison, 2009; Ponitz, McClelland, Jewkes, Conner, Farris & Morrison, 2008) in children's early years. HTKS requires children to remember and respond to behavioral commands. It has predictive validity with achievement and teacher-ratings of self-regulation.

vii. **Children's health status:** In line with similar international studies (Fernand, Gertler & Neufeld, 2008; Overholt, Sellers, Mora, Paredes & Herrera, 1982; Walker, Wachs, Meeks Gardner, et al., 2004) we collect information on height and weight, BMI and arm circumference once a year following World Health Organization (WHO) standards.

Data Collection

Instruments were chosen based on recommendations for child development research in low-income countries and used in previous evaluations in these contexts (Fernald, Kariger, Engle and Raikes, 2009). Most instruments have been used extensively in evaluations of early care or education. We have chosen instruments for the younger ages are that are highly predictive of the instruments chosen for later ages. We use a combination of instruments that are good measures for each age group with measures of less depth but longer time spans. Measures have demonstrated adequate psychometrics in similar longitudinal studies with Latin American or other Hispanic populations and have effectively assessed program effects in other Latin American evaluations. Table 2 summarizes the instruments used by area and age.

Table 2. Evaluation Instruments by Child's Age

Measurement Instruments	AGE OF CHILD at APPLICATION							
	0-1 yr	1-2 yrs	2-3 yrs	3-4 yrs	4-5 yrs	5-6 yrs	6-7 yrs	7-8 yrs
Moderators & Controls								
1 Parent Survey	X	X	X	X	X	X	X	X
Home Patterns								
2 HOME	X	X	X	X	X	X	X	X
Health								
3 Anthropometrics	X	X	X	X	X	X	X	X
4 Immunization Records	X	X	X	X	X	X	X	X
5 Meal Deprivations	X	X	X	X	X	X	X	X
Socio Emotional								
4 Socio-emotional Scale: ASQ: SE	X	X	X	X	X			
5 BASC					BASC-P	BASC-C	BASC-C	BASC-C
Non-verbal cognitive development								
6 Executive Function: HTKS				X	X	X	X	
Language and Mathematics Development								
7 Bayley (0-3 yrs only): Cognitive Scale	X	X	X					
8 Bayley (0-3 yrs only): Language Scale	X	X	X					
9 TVIP (2.5 and up)			X	X	X	X	X	X
10 ELSA (3 and up)				X	X	X	X	
11 WM-R (3 and up)				X	X	X	1, 9, 10	1, 9, 10
12 Vineland (3 and up)				X	X	X	X	X
13 School status	X	X	X	X	X	X	X	X
Motor Development								
14 Bayley (0-3 yrs only): Motor Development	X	X	X					

Each data collection takes two months per community. We have kept a group of 5 graduated psychologists since 2010 for data collection. This team was trained and certified for reliability by NIEER personnel in 2010 in child assessments and classroom ratings. Every year since then, NIEER has prepared a refresher for the team in both instruments emphasizing specific issues found in data entry for the previous year, ensuring sustained reliability of the assessors. The unique opportunity of keeping the same team for over three years has enriched the field work experience immensely. People in the communities are well acquainted with our group and are cooperative at arrival for evaluation.

Upon arrival in the community, we rent a house as close as possible to the center and adapt it for child assessments and parental interviews. This way, we guarantee that children are assessed in the same environment and under the same conditions. Parents are called in advance and given an appointment to attend our evaluation house. Daily appointments are well organized and teams of psychologists and interviewers collaborate to carry out both child assessments and parental interviews.

Each child and his/her parent takes between one hour and an hour and a half depending on his/her age. All data is collected in paper in field. After the file of each child is completed, it is revised by the group manager for consistency and sent from Santa Marta to Bogotá (Colombia's capital city), where the headquarters of the field work provider is located. At this point there is a second round of supervision. The collected data and a report are sent to the research team at Universidad de los Andes on a weekly basis. At the university, a third round of supervision takes place for completeness, consistency and appropriate data reporting. The files are fully copied. Originals are sent via physical mail to NIEER where data entry takes place and copies are stored in Bogotá for at least one year.

Power

Randomization of household ensures comparability of treatment and control groups and sufficient power at the lowest cost for a longitudinal study as it minimizes the required sample size. The relationship between power and sample size was estimated for effect sizes of 0.30sd and of 0.25sd (average effects for continuous outcomes in verbal ability and/or cognition for programs of this type in the region; see Nores and Barnett, 2010) and a randomized design with repeated measures where children are measured annually for five periods (non-clustered).

Power calculations for our research design were conducted using the Optimal Design for Longitudinal and Multi-Level Research software (University of Michigan, 2006). We expected effect sizes of .30 or more given the intensity of the intervention, so that this sample size allows for the investigation of subgroup differences at individual ages. Power analyses indicated a power of 1 with a sample size of 1200 children (and above .85 for sample sizes of 700, which would accommodate attrition or compliance issues and allow for subgroup analyses for infants and toddlers versus preschoolers). Power analyses were conducted with statistical significance level set at $\alpha=0.05$, using a two-tailed test, F=1 (assessments per year), D=5 (years of data) and effect

sizes set at 0.30sd & 0.25sd, respectively (Figure 1 below). A refresher may be necessary if higher attrition occurs among younger children in subsequent years, but this has not proved necessary so far. We randomized additional children that entered the program in subsequent years to allow the possibility of using such children in a refreshing of the sample if this becomes necessary.

Figure 1. Power Analyses
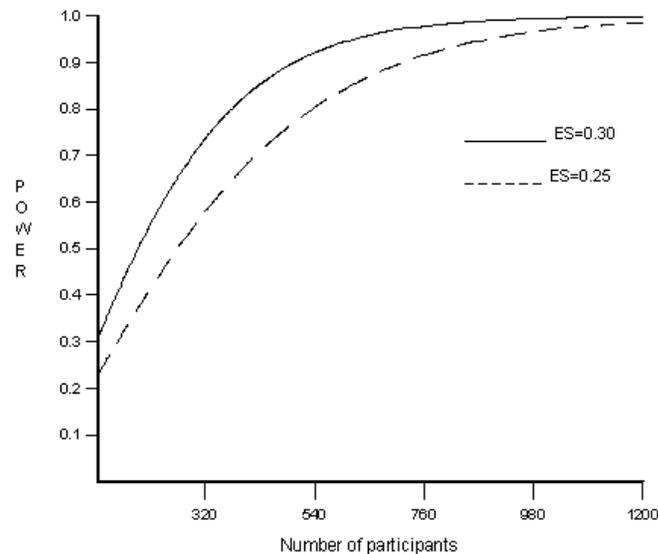
Follow-ups and Reducing Attrition Risks

Table 1 presents our study's sample size by age, group and intent to treat status at baseline (year 2010). A total of 1,219 children were interviewed in 2010 prior to the introduction of the intervention, 544 in Timayui and 675 in La Paz. These same children have been followed since in 2011, 2012 and currently in 2013. In 2011 and 2012 we had an average 5% attrition rate with respect to baseline, which we consider a low rate given the characteristics of these communities. Approximately 5% of children in the sample have been assessed in other places (communities and cities) different from the two original communities in Santa Marta.

Table 1. Description of the study sample (baseline 2010)

Age group	Timayui		La Paz		Total	
	Lottery winners	Lottery losers	Lottery winners	Lottery losers	Lottery winners	Lottery losers
<1	38	75	30	24	68	99
1-2	63	80	56	96	119	176
2-3	82	61	56	123	138	184
3-4	57	45	64	97	121	142
4-5	29	14	69	60	98	74
Total	269	275	275	400	544	675

By 2011 we interviewed only 20 children younger than 1 year of age and approximately 200 older than 5 years of age as they grow with the study. By 2012 we had around 300 children already in primary school. In 2011, children had been exposed for at most 8 months to the program. By 2013 children will have been exposed to the program for at most 3 years.

Attrition rates have been kept low (less than 10 percent) through a variety of strategies that include:

i.   **Call center:** a study call center calls each family at least once every two months, sends happy birthday messages and receives calls for questions regarding the study;

ii. **Monetary incentives:** these are delivered to families after each child's evaluation is completed every year in the amount of approximately USD 10;

iii. **Comprehensive contact directory:** created through parental interviews and through the work of field work teams, with information on both parents and on friends and other family of the child; and

iv. **Close work with the communities:** the field work team visits children wherever they have moved to or pays for travel expenses for families to visit our assessment hubs in the communities.

Analyses

Statistical analyses are performed to address these questions:

1. What are the immediate and short-term effects of enrollment and participation in aeioTu on children's early learning and development?
2. To what extent do these effects vary by intensity (cohorts)?
3. Do pre-school effects persist over time, and in particular as children transition to primary?
4. Are pre-school effects observable in areas beyond school outcomes?
5. To what extent do child and family characteristics (gender, ethnicity, parents' education) and location moderate the effects on children? (Additional exploratory analyses can expand the list investigated)
6. To what extent do effects differ with variations in implementation and classroom experiences?; and
7. What is the cost-effectiveness of this intervention?

Tabulations, linear and random effects models and Instrumental Variable models will be used to estimate the impact of aeioTu centers on children by intent-to-treat and by treatment-on-treated. Investigation of mediators will employ growth models and structural models of relationships over time.

Tabulations for outcomes and covariates by intent-to-treat (ITT) status: The treatment group is defined as the group of lottery winners (intended to be treated) and the control group is understood as the group of lottery losers. While there were differences between intent-to-treat

and actual enrollment (treatment), we will report outcomes for the enrolled versus non-enrolled groups in a complementary report since these require statistical controls. ITT analyses of the outcomes of an experiment are based on the initial lottery (randomization) and not on who did in fact receive treatment. Randomized trials analyzed through the ITT approach provide unbiased comparisons between of the outcomes of the two groups (winners and losers of the lottery). Since in effect there is always some degree of non-compliance on the random assignment (some children who won the lottery never attend the program, and some in the control group end up attending the program) ITT is preferred since it gets around these crossover (or dropout) problems. This means that the results should be interpreted as effects of the availability of the aeioTu program rather than as the effects of attending the centers.

Linear Models: Given the design of the experiment, the ensuing analysis will be relatively straightforward. We will first estimate the following model:

(1) $$A_i^t = \lambda_1 + \lambda_2 ITT_i + \lambda_3 A_i^{baseline} + \lambda_4 X_i + \varepsilon_i$$

Where any of the outcome measures of child i from the follow-up tests and survey and ECDi is an indicator variable that equals one if the child belongs to the center group. As the children who attend center will be randomly selected, we can be sure that the exclusion restriction holds (i.e. $E(ITT_i, \varepsilon_i)=0$). Thus, λ_2 will capture the causal impact of belonging to the ITT group on the outcome.

Equation 2 provides the reduced-form relationship between the program existence and outcomes. Measuring this effect is important because it informs us of the policy impact of belonging to the ECD group versus the control group, thus allowing us to calculate the relative benefits and costs of the two groups. Within this model, we will follow Barron and Kenny's (Baron and Kenny, 1986) four-step procedure for establishing mediation and estimate and plot interactions to interpret effects in a moderator model (Kelley and Maxwell, 2010). All assumptions will be cross-checked visually.

Variations of this model will inquire into heterogeneous effects by age groups (infants and toddlers or preschoolers) and by initial developmental levels and also include inquiry into family fixed effects as the program was randomized by families, rather than by children.

Instrumental Variable (IV) Models: IV models or two-stage least squares estimation models will be used to analyzed the impact of treatment-on-treated (TOT). These models adjust for non-compliance on the random assignment (some children who won the lottery never attend the program, and some in the control group end up attending the program). In essence, the models estimate the relationship between program treatment and outcomes, instrumenting program treatment based on the lottery outcome.

(2) $\quad A_i^t = \lambda_1 + \lambda_2 Treatment_i + \lambda_3 A_i^{baseline} + \lambda_4 X_i + \varepsilon_i$

(3) $\quad Treatment_i^t = \lambda_1 + \lambda_2 ITT_i + \lambda_3 A_i^{baseline} + \lambda_4 X_i + \varepsilon_i$

Additional models such as growth curve analysis and structural equation modeling will be considered (Barnett, Young and Schweinhart, 1998; Robin, Frede and Barnett, 2006).