Improving Access and Quality in Early Childhood Development Programs: Experimental Evidence from The Gambia*

Moussa P. Blimpo
World Bank

Pedro Carneiro
University College London
Institute for Fiscal Studies
Centre for Microdata Methods and Practice

Pamela Jervis
Institute for Fiscal Studies

Todd Pugatch
Oregon State University
IZA

June 2019

Abstract: This paper studies two experiments of early childhood development programs in The Gambia: one increasing access to services, and another improving service quality. In the first experiment, new community-based early childhood development (ECD) centers were introduced to randomly chosen villages that had no pre-existing structured ECD services. In the second experiment, a randomly assigned subset of existing ECD centers received intensive provider training. We find no evidence that either intervention improved average levels of child development. Exploratory analysis suggests that, in fact, the first experiment, which increased access to relatively low quality ECD services, led to declines in child development among children from less disadvantaged households. Evidence supports that these households may have been steered away from better quality early childhood settings in their homes.

JEL Codes: I25, I38, O15, O22

Keywords: early childhood development; cognitive stimulation; teacher training; The Gambia; randomized control trials; Malawi Developmental Assessment Tool

* Author contact information: mblimpo@worldbank.org, p.carneiro@ucl.ac.uk, pamela_j@ifs.org.uk, todd.pugatch@oregonstate.edu. We thank the World Bank and the Japan Social Development Fund (JSDF) and the Strategic Impact Evaluation Fund (SIEF) for financial support. Pedro Carneiro gratefully acknowledges the support of the ESRC for CEMMAP (ES/P008909/1) and the European Research Council through grant ERC-2015-CoG-692349. We thank the Gambian Ministry of Basic and Secondary Education (MoBSE) for their active support and engagement; the National Nutrition Agency (NaNA) for the Implementation of the intervention; The Gambia Bureau of Statistics (GBOS) for organizing the data collection. Finally, we thank Yi-Kyoung Lee, Elizabeth Joof, Hervé Akinocho**,** for their guidance and support for our research. Roberta Nilson, Ruifan Shi, and Yabin Xu provided excellent research assistance. All errors are ours. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

1. Introduction

Early childhood experiences lay the foundation for outcomes later in life (see Currie and Almond 2011 for a review of the U.S. literature; Nores and Barnett 2010 for evidence outside the U.S.; and Tanner, Candland, and Odden 2015 for developing countries). A key mechanism underlying this link is the rapid development of the brain during the preschool years. During this period, children develop their abilities more fully in stimulating environments. But in developing countries, inadequate cognitive stimulation has been identified as one of four key risk factors hindering childhood development (Walker et al. 2007, 2011).[1] In these countries, 250 million children under age 5 risk failure of reaching their development potential (Black et al. 2017).

More than 80 million of these children live in Sub-Saharan Africa, representing two-thirds of children in the region. Countries in Sub-Saharan Africa face two main challenges to promoting children's development. The first challenge is lack of access to early childhood development (ECD) services. Enrollment in pre-primary schooling is 22% in Sub-Saharan Africa, compared to 79% in OECD countries (World Bank 2015). The second challenge is improving the quality of existing ECD services. Even where access exists, classroom quality can be low and uneven due to resource constraints and low skills of ECD providers.

This paper evaluates two experiments to address these challenges in one of the poorest countries in the world, The Gambia (per capita income 427 USD; World Bank 2016). The experiments tested alternative approaches to deliver a new curriculum intended to stimulate development of children aged 3-6 through structured play. In the first experiment, new community-based early childhood development centers were introduced to randomly chosen villages that had no pre-existing structured ECD services. This intervention increases access to new ECD services.

In the second experiment, a randomly assigned subset of existing ECD centers operating as annexes in primary schools received intensive provider training to implement the new curriculum. A control group of ECD centers (also annexes) received the new curriculum without the provider training. This intervention improves the quality of existing ECD services.

[1] The other risk factors are stunting, iodine deficiency, and iron deficiency anemia.

We find no evidence that either experimental intervention improved average levels of fine motor skills or language and hearing, our key child development measures. We also find no evidence of effects at other points in the outcome distribution.

Despite these results for the full sample, the community-based ECD intervention had differential effects according to several baseline characteristics. Exploratory analysis suggests that children from more advantaged households scored about 0.4 standard deviations lower in language skills when living in a community randomly assigned to the community-based ECD treatment. We find suggestive evidence that these households may have been steered away from better quality early childhood settings for their children in their homes.[2] Nevertheless, parents in treatment communities still send their children to these facilities either because they have an erroneous perception of classroom quality, or because of the value of the time freed up by sending their children to preschool instead of caring for them at home, though we lack evidence to distinguish these possibilities. In contrast, we find no differential effects of the provider training treatment in ECD Annex centers according to children's background.

The type of centers in each experiment, as well as the nature of the interventions, are very different. Community-based centers are provided when no ECD services exist in the community. There are few or no physical or human resources to build on, which means that the services provided are of limited quality. New but very basic infrastructure needs to be built to house the centers. Teachers are residents in the community with few qualifications and experience. In the type of communities where no centers exist, these community-based centers are made possible by leveraging minimal investments and local human resources. Currently, they are the only center-based alternative that the government can provide at scale when no ECD services exist in the community.

In contrast, ECD Annexes operate in the existing infrastructure of primary schools. This strategy allows for teachers with much higher qualifications and experience than in communities where no ECD services exist. Like community-based centers, ECD Annexes also leverage local physical and human resources to provide services, but these resources are likely of higher quality

[2] Negative impacts of low quality ECD have been found elsewhere (see, e.g., Baker, Gruber, and Milligan 2008; S. Berlinski and Schady 2015; Ichino, Fort, and Zanella 2019; Bernal et al. 2019).

in communities with an existing primary school than those with a community-based center. In fact, our own baseline evidence suggests this is precisely the case.

The first intervention therefore improves access to community-based ECD services, from a situation of no access whatsoever to one where minimal quality ECD centers are provided. In contrast, the second intervention aims to increase the quality of already existing ECD centers in school annexes, which are of higher quality than ECD community-based centers even in the absence of treatment.

In sum, we find no evidence that the minimal quality centers that are provided in these settings are helpful for child development. If anything, the opposite is true for children in better home environments. Our findings of no effects for provider training in existing ECD centers suggest that any quality improvements were insufficient to improve child outcomes.

A central aspect of our paper is that, in both experiments, we consider public child care provision designed to be delivered at scale, with the available physical and human resources in very challenging settings (as opposed to, for example, small pilot programs implemented by very specialized and highly qualified staff). The ECD centers we study are currently provided by the Gambian government at national scale.[3] They are similar to the types of ECD centers that can be delivered by the poorest governments in Sub-Saharan Africa.

We make two main contributions to the literature on interventions for preschool children in developing countries (for a review, see J. Behrman, Fernald, and Engle 2013). First, we provide evidence on the effects of preschool access from a randomized control trial of a national-scale program in Sub-Saharan Africa. This combination of research design, program delivery, and geographic focus is rare in an otherwise growing literature on the effect of access to ECD programs.[4] The studies most closely related to ours consider programs that operate at national or otherwise large scale (e.g., Berlinski and Galiani 2007; Berlinski, Galiani, and Manacorda 2008; Berlinski, Galiani, and Gertler 2009; Rosero and Oosterbeek 2011; Bastos, Bottan, and Cristia

[3] The experiments we analyze covered only two of six regions, however. We provide more details on program design in the following section.

[4] For example, J. R. Behrman, Cheng, and Todd 2004 for Bolivia; Aboud 2006 for Bangladesh; Berlinski and Galiani 2007, Berlinski, Galiani, and Gertler 2009 for Argentina; Berlinski, Galiani, and Manacorda 2008 for Uruguay; José Rosero and Oosterbeek 2011, Jose Rosero 2012 for Ecuador; Rao et al. 2012, Bouguen et al. 2018 for Cambodia; Martinez, Naudeau, and Pereira 2013 for Mozambique; Krafft 2015 for Egypt; Attanasio et al, 2018, and Pinto, Santos, and Guimarães 2016 for Brazil; Bastos, Bottan, and Cristia 2017 for Guatemala).

2017; and Bouguen et al. 2018). However, there are few randomized control trials of expanded access to ECD in developing countries, such as Martinez, Naudeau, and Pereira (2013), Bouguen et al. (2018), and Bernal et al. (2019). Of these, only Martinez, Naudeau, and Pereira (2013) study a program in Sub-Saharan Africa.[5]

Most of the aforementioned studies find positive effects on children's development and subsequent school performance, with some exceptions. Consistent with evidence from developed countries, the counterfactual environment for children in ECD programs is often decisive for impact estimates. Where the alternative to ECD programs is a lower quality home environment, estimated effects will be positive. Alternately, when ECD participation substitutes for other forms of schooling, such as early primary school attendance (Bouguen et al. 2018), or reduces parenting quality by increasing maternal labor force attachment (Rosero and Oosterbeek 2011; Rosero 2012), impact estimates can be zero or negative. These impacts can also vary within a program according to household characteristics (Pinto, Santos, and Guimarães 2016).

Our second contribution is to expand the relatively small literature on improving the quality of existing ECD programs via teacher training.[6] Despite a few studies using credible identification strategies (Baker-Henningham et al. 2012 for Jamaica, Yoshikawa et al. 2015 and Bowne et al. 2016 for Chile, Bernal 2015 for Colombia, and Araujo et al. 2018 for Peru), there is scant and mixed evidence on provider training in Sub-Saharan Africa. Ozler et al. (2016) find that ECD providers in Malawi who were randomly assigned to receive in-service training improved classroom organization and teaching quality, but these behavioral changes increased child development only when paired with parent education. Wolf, Aber, and Behrman (2017) evaluate an experiment with a similar design in Ghana, but find gains in child development only when provider training was *not* paired with parent education. Additionally, even when programs to improve quality are effective in RCTs, gains may not be sustained when taken to scale. We build on the important contributions of these studies, adding to the thin evidence base on ECD provider training in Sub-Saharan Africa. Importantly, in our setting the programs are

[5] Several other studies have evaluated pre-school attendance in Sub-Saharan Africa (Taiwo and Tyolo 2002 for Botswana; Mwaura, Sylva, and Malmberg 2008; Malmberg, Mwaura, and Sylva 2011 for Kenya, Tanzania, and Uganda; Zuilkowski et al. 2012 for Zambia), but these studies lack credible identification strategies to distinguish the role of ECD exposure from unobserved child or parent attributes.

[6] Studies of efforts other than provider training intended to upgrade ECD quality in developing countries include Armecin et al. (2006) for the Philippines; He, Linden, and MacLeod (2009) for India; and the many follow-ups to the Mauritius Child Health Project (Raine, Venables, and Mednick 1997; Raine et al. 2001, 2003, 2010).

implemented at scale by public providers, within the same context, providing evidence that is immediately relevant for policy.

In the next section, we describe the program and data. Section 3 describes our empirical methodology. Section 4 presents experimental results. Section 5 concludes.

2. Program Description and Data

The Gambia is a small West African country with population 1.9 million and per capita annual income of 427 USD (World Bank 2016), making it one the 10 poorest countries in the world (World Bank 2015). Its education system is divided into six numbered administrative regions. Region 1 is the capital, Banjul, on the Atlantic coast, with Regions 2-6 located in increasingly remote areas toward the East. Regions become more rural and poor as distance from the capital increases.

Other than informal home care, ECD services exist in three forms in The Gambia: 1) private centers, located mostly in relatively urban areas and serving richer children; 2) public centers which are built as annexes to primary schools (hereafter referred to as ECD Annexes); and 3) community-based centers, which are publicly run, stand-alone facilities located in communities without primary schools. ECD Annexes and community-based centers do not charge fees. The experiments in this study are part of major government efforts to increase ECD access and quality.

ECD access has expanded greatly in recent years, in line with government's goal of integrating ECD programs into the standard primary school sequence. Gross enrollment in ECD programs grew from 22% in 2007 to 37% in 2013 (Zoyem 2010; Gambia Ministry of Basic and Secondary Education 2017). However, there are considerable regional disparities in ECD enrollment rates, ranging from a low of 20% (Region 5) to a high of 45% (Region 2; Gambia Ministry of Basic and Secondary Education 2017). These differences mirror similar disparities by family income, implying that a large population of Gambian preschool children are excluded from center-based early development and learning opportunities, especially those from rural and poor households.

To improve quality, the Gambian Ministry of Basic and Secondary Education (MoBSE, hereafter "the ministry") implemented a new early childhood curriculum in 2012. The new

curriculum, known as Gambia Open and Active Learning Spaces (GOALS), aims to promote creativity, problem solving, and confidence through structured play. Each day includes time for group activities, games to promote critical thinking, physical development, and music/singing. The curriculum runs for 40 weeks annually, concurrent with the academic calendar. Activities run for four hours daily Monday through Thursday and three hours on Friday. Teachers are expected to spend an additional 1.5 hours each day to prepare for the next day's session. Additionally, a monthly meeting is held with parents to discuss the program and children's progress.

The curriculum was deployed at all ECD Annexes and community-based centers nationally, replacing *ad hoc* local approaches with a common standard. To assess the effectiveness of different approaches to curricular delivery, the ministry experimentally implemented two interventions in parallel.[7] Both experiments occurred in Regions 2 and 6, with treatments assigned at the village level. One intervention aimed to increase access to ECD center-based care, while the other aimed to improve the quality of existing center-based care.

In the first experiment, to increase ECD access, the ministry built new community-based centers in randomly chosen villages that had no pre-existing structured ECD services. Each community-based center delivered the new GOALS curriculum. Management committees, comprised of parents and community leaders, were formed to oversee the centers. Most care providers recruited to these centers were community members with no previous experience and less than a high school degree. We refer to this as the community-based treatment, and the corresponding control group of eligible villages without ECD services as the pure control.

In the second experiment, a teacher training program was delivered to a randomly chosen subset of existing ECD Annex centers. Teachers received intensive training in the new curriculum in three sessions between September 2012-September 2013. The trainings lasted five, eight, and eight days, respectively. The Gambia office of ChildFund, an international NGO, conducted the trainings, which were financed by the Japan Social Development Fund through the World Bank. We refer to the experimental groups in this case as ECD Annex treated and ECD

[7] A third experiment, on a different ECD program. was conducted for children ages 0-3 (Blimpo, Carneiro, Jervis and Pugatch, 2018), who were too young to be eligible for the services studied in this paper. This experiment occurred in a separate set of communities than the experiments described here.

Annex control. The ECD Annex control group received no training on the new curriculum, as there are no regular opportunities for in-service teacher training in The Gambia.

Teachers in the community-based ECD centers of the first experiment also received the training provided to the ECD Annex treatment group. Compared to the pure control group, the community-based treatment therefore represents an increase in access to new ECD services, including site construction, formation of management committees, curriculum, and teacher training. All ECD Annexes in the program regions received the new curriculum, but only the ECD Annex treatment group received the teacher training. The ECD Annex treatment therefore represents an attempt to improve quality through teacher training, since the same curriculum was delivered to all annexes. Table 1 summarizes the research design.

Figure 1 shows a map of sites included in the sample. The bulk of the sample sites are in Region 6, one of the poorest and the most remote regions of the country.[8] Treatment was stratified by region to ensure sample balance.[9] The sample for the ECD Annex experiment includes 26 treatment sites and 27 control, while the sample for the community-based ECD experiment includes 40 treatment sites and 51 control.[10]

The new curriculum and initial teacher training began in September 2012, the start of the 2012-2013 academic year. A baseline survey was conducted before the beginning of the academic year, in May-July 2012.

[8] Table SA1 of the Supplemental Appendix uses the 2003 Census, the most recent conducted before the program began, to compare communities in program Regions 2 and 6 with the rest of the country (Regions 1 and 3-5). Regions 2 and 6 differ from each other on many dimensions, with Region 2 more populous, educated, and developed. These differences reflect the more urban character of Region 2 compared with rural and remote Region 6. Stratifying the sample by region ensures that these differences are not spuriously correlated with treatment. Regions 2 and 6 also differ from the rest of the country, as shown by the many significant differences reported in columns (4)-(5) in the table. These differences suggest that the treatment effects reported in this study may not generalize to all regions of the country. However, the heterogeneity between Regions 2 and 6 ensures that the program occurred in a broad range of contexts found within the Gambia.

[9] We include a dummy variable for Region 2 to account for this stratification throughout all analyses.

[10] The sample of ECD Annexes is smaller than the community-based ECD experiment because it the latter forms an exhaustive list of ECDs in the two regions. A larger number of control sites were sampled in the community-based ECD experiment because these were also used as control sites for a separate experiment with children ages 0-3 (M. P. Blimpo et al. 2018), allowing for economies of scale in data collection. We dropped one ECD Annex treatment site because none of the sampled children met the age eligibility criteria according to their birth certificates. There are two fewer community-based ECD and pure control sites in the baseline data than in the full sample for the same reason.

In the community-based ECD experiment, the reference population was children aged 3-6 in the sampled communities. After making a complete listing of all households with children in that age group, a random sample of 16 households was taken from each community in the treatment and control groups. In households with multiple eligible children, one was randomly sampled. In the ECD Annex experiment, since these are existing centers with enrollment rosters, 16 children (and thus their households) were randomly selected from the roster to participate in the study sample.[11]

For both experiments, the household head and main caregiver for the eligible child were interviewed, with modules on household assets, expenditures, employment, demand for ECD services, health (own and the child's), parenting activities and home environment, and attitudes towards disciplining the child. Anthropometric measurements (height, weight, brachial circumference) were taken from children.

An endline survey was conducted in November-December 2013, 14 to 15 months after treatment began. The endline survey asked similar questions as the baseline, with an additional module to assess caregiver knowledge of childhood nutrition and health. Endline participants included baseline households and newly randomly sampled households, allowing for an increase in the sample size of analyses using only endline data. We analyze attrition from the survey in the Data section. Figure 2 shows a timeline of the project and research milestones.

In addition to the survey modules previously mentioned, children in sampled households were given the Malawi Developmental Assessment Tool (MDAT), a test of child development designed for rural Africa (Gladstone et al. 2010). The test consists of two modules: 1) fine motor skills, and 2) language and hearing.[12] Each module has multiple versions tailored to different age ranges. The fine motor skills tool asks children to complete tasks such as stacking blocks in various configurations, placing pegs in a board, and determining the relative weight of objects. The language and hearing test requires children to point to body parts by name, identify the names or uses of objects, identify the letters in one's name, and similar tasks. Each item on the tests is marked as complete or incomplete, with the overall score determined by the total number

[11] A listing of eligible households was not conducted for the ECD Annex experiment, as such, we do not have statistics on the extent of uptake of the ECD Annexes where they exist. The reference population is therefore children in the relevant age group who were enrolled in the ECD annexes at the time of the study.
[12] Assessments were conducted in the mother tongue or English, whichever was more comfortable for the child.

of completed items. Enumerators received five days of training on the tests before administering them in the field. Due to interviewer time constraints, enumerators were instructed to administer the test to children in a random subsample of surveyed households. Children took the MDAT at home regardless of treatment status.

The baseline and endline used the MDAT versions intended for children aged 36-59 and 53-76 months, respectively (though children completed the tests regardless of their age at the time of the surveys).[13] We assessed the reliability of the MDAT in our sample by calculating Cronbach's α separately by module (fine motor and language and hearing) and survey wave (baseline and endline), using all available MDAT scores in the survey (including children from the pure control group and community-based ECD treatment). The Cronbach α values ranged from .82 to .88, indicating high reliability to measure the underlying constructs. We also check internal consistency of MDAT scores in our sample by examining their relationship with child age and household wealth. Figure 3(a) shows that both fine motor and language skills progress approximately linearly with age. Figure 3(b) plots MDAT scores against deciles of a wealth index, with scores generally increasing in wealth, particularly for the top deciles. The figure also shows the progression of height-for-age with wealth as a comparison.[14]

Program implementation, though largely successful, encountered significant challenges. 14 out of 40 community-based ECD treatment sites reported implementation problems. The most common problems reported were absent or sick teachers or lack of materials. Two sites reported that their facility had not been constructed, forcing teachers to provide services outdoors. Because of administrative issues, many teachers went several months without receiving their stipends leading to some resignations and replacements. No ECD Annex treatment sites reported implementation problems. We provide additional details on implementation in Section 4.1.

3. Methodology

[13] Supplemental Appendix A presents the MDAT versions used in the surveys.

[14] Figure 3(b) plots our preferred measure of MDAT scores used throughout the paper. We adjust MDAT scores for age by regressing a child's raw score on age (in months), age squared, and a female dummy, then dividing the residual by its standard deviation. We run this procedure separately for the fine motor and language and hearing modules and for baseline and endline. We calculate height-for-age using the international benchmarks defined by the World Health Organization. The wealth index is the first principal component of household asset ownership. Figure 3(a) does not adjust MDAT scores by age.

We can analyze the effect of each experimental treatment through a comparison of mean outcomes between children in treated and control communities. We cluster all standard errors at the community level to adjust for correlated outcomes among units exposed to the same treatment. To test for differences between treatment and control groups, we further adjust for the stratification of treatment status within regions. We make this adjustment by regressing the outcome on an indicator for treatment and a dummy for whether the community is in Region 2:

$$y_{ic} = \alpha + \beta D_c + \gamma Region2_c + \varepsilon_{ic} \quad (1)$$

where *i* indexes the child, *c* indexes the community, *y* is the outcome, *D* and *Region2* are indicators for treatment and Region 2, respectively, and ε is an error term. We estimate this equation separately for the community-based ECD and ECD Annex experiments, with the definition of treatment changing accordingly. In each case, the coefficient of interest is β, which measures the difference in mean outcomes between children exposed to each treatment compared to the corresponding control group. We will also run versions of equation (1) that include the baseline outcome y_{0ic} on the right-hand side in order to improve the precision of our estimates.

Our main outcomes of interest are scores for the two MDAT modules, fine motor skills and language and hearing. We normalize each score by child gender and age, using children from both experiments as the underlying population (outcomes are then measured in standard deviations). Additionally, we analyze subsets of MDAT items which are most closely aligned with the structured play format of the GOALS curriculum, such as stacking and counting blocks. For each subset, we measure the percentage of items successfully completed by the child.

The parameters estimated by equation (1) are intent to treat (ITT) estimates. We discuss the relationship between our ITT estimates and the local average treatment effect (LATE) in Section 4.1. For the ECD community center experiment they are likely to be a good approximation to the average treatment effect since there are little to no reports of eligible households refusing enrolment in these centers (as we discuss below, this comes from indirect reports from program staff, since we have not collected this information in our surveys). For the ECD Annex experiment, we do not expect the treatment to affect enrolment rates, which are already very high, bordering 90%. Therefore, a good approximation to the average treatment effect of teacher training for children attending ECD Annexes is (approximately) $\beta/0.9$.

In Appendix B, we also show how we conduct non-experimental comparisons of the effectiveness of ECD Community Centers versus that of ECD Annexes. This is important because our two experiments concern these two potentially very different types of centers. In fact, as we will see in the next section, teachers in Annexes have much more experience and education than those in Community Centers. Consistent with this, Appendix B documents that Annexes are of higher quality than Community Centers, even in the absence of the teacher training treatment.

4. Experimental results

4.1. Attrition and baseline balance

Table 2 presents sample sizes and analyzes attrition. The sample includes all children aged 3-6 at the time of the baseline. Panel A, columns (1)-(2) show the number of children sampled in baseline and endline, respectively, separately by experiment and treatment group. The endline sample is split into two categories: those who appeared in the original (baseline) sample, and those newly sampled to increase the sample size. Columns (3)-(4) of Panel A present the same information but restrict the sample to those who completed the MDAT fine motor skills and language and hearing tests. Some children present in the baseline survey completed the MDAT at the endline but not the baseline, and therefore appear in the "original sample" group of column (4) although they lack a baseline score. For this reason, in column (5) we show the number of children who completed the MDAT in both baseline and endline.

This table shows that, for both experiments, only about half the children in each survey wave are administered the test, as explained in Section 2. The set of households where testing was conducted was selected at random in both survey waves.

Nevertheless, there are some imbalances in the characteristics of test takers and non-test takers, which occur purely by chance. At baseline, test completers come from statistically significantly larger households, are more likely to have a household head in agriculture, and have higher vaccination rates than non-completers in the community-based ECD experiment. In the ECD Annex experiment, test completers are younger, have more educated parents, and higher vaccination rates than non-completers. These differences should be kept in mind when interpreting our results, which apply only to children completing the test. Fortunately, test completion is uncorrelated with treatment in either experiment, increasing confidence in the

internal validity of our estimates.[15] We later check robustness of results when using multiple imputation to infer the outcomes of children who did not take the MDAT.

Panel B of Table 2 analyzes sample attrition. We define two types of attrition: 1) attrition from the survey, in which a household that completes the baseline survey fails to complete the endline survey, and 2) attrition from the test, in which a child that completes the baseline MDAT does not complete the endline MDAT. Columns (1)-(3) of Panel B show the sample proportion of each type of attrition by treatment status. Attrition in the community-based ECD experiment exceeds that of the ECD Annex experiment. Within each experiment, however, neither type of attrition differs significantly between treatment and control groups, as shown by the p-values in column (5). In what follows, we restrict attention to children who complete the MDAT, unless otherwise indicated.

Table 3 presents baseline characteristics and tests for balance between treatment and control groups within each experiment. Columns (1)-(4) show the control group mean, treatment group mean, difference, and corresponding p-value for the community-based ECD experiment. Columns (5)-(8) repeat the exercise for the ECD Annex experiment. Baseline values across all groups demonstrate the economic and social disadvantages faced by The Gambian children. Height-for-age, which proxies for early life nutrition, is more than 1.25 standard deviations below the international average for all groups. Average completed schooling of mothers is 2 years or less. Mean household expenditure per capita ranges from US$418-523. 13-18% of children were ill at the time of the survey, and mothers report high levels of mental distress (measured as the percentage of indicators of distress, such as feeling lonely, sad, or fearful, experienced “most of the time”).

Looking across experiments, children in the community-based ECD experiment are younger and score lower in fine motor skills, language and hearing, and height-for-age than those in the ECD Annex experiment. These disadvantages are not surprising as the presence of ECD Annexes in these communities may suggest greater opportunities for investment in children. Notably, household socioeconomic indicators are not uniformly higher in the ECD Annex experiment communities. Additionally, the primary schools (and thus the annexes) are not always close to all communities they serve. In an environment where children walk themselves

[15] We report these results in Tables SA2-SA3 of the Supplemental Appendix.

to schools, younger children (closer to 3) are unlikely to be enrolled in annexes relative to older children (closer to 6). The community-based version, however, creates the conditions for access to all children within the age range.

Within experiments, treatment and control groups are broadly similar. Fine motor skills, language and hearing, and height-for-age are not significantly different between treatment and control in either experiment. Differences in most other characteristics are also not statistically significant, with a few exceptions. Children in the control group of the community-based ECD experiment are 14 percentage points more likely than the treatment group to be attending an ECD program at baseline, significant at 1%. There are also a few differences that are significant at the 10% level, including a greater proportion of treated children with a household head in agriculture and lower assets. In the ECD Annex experiment, treated children have lower household expenditure than control children, significant at 10%.

Despite these imbalances, we have no reason to believe that randomization was compromised. The randomization procedure was carefully monitored by the research team and compliance with randomization was perfect. Instead, we attribute baseline differences to chance. Controlling for baseline MDAT scores, which summarize the cumulative effect of early childhood investments, should mitigate spurious treatment effects estimates. Additionally, we estimated treatment effects controlling for the variables imbalanced at baseline, and find similar results. We discuss these results in greater detail below.

4.2. Take-up and implementation

Take-up rates of ECD services were not recorded in the endline survey. They are however quite high. In villages in the community-based ECD experiment, there were no reports of households refusing the services nor of waitlists. At baseline, most sampled children in these villages were cared for at home. The 30% already attending ECD (see Table 3 for breakdown of treatment vs. control) suggests high latent demand for ECD services. Indeed, 15% of the sample reports attending an ECD Annex or community-based ECD center at baseline, which by definition is located outside the village.[16] A simple calculation provides a way to address this. Suppose we take the 90% baseline ECD attendance in ECD Annex villages as an approximation

[16] The remaining children attend private ECD (5%), madrassas (12%), or other (2%). Totals sum to more than 30% because households could report more than one type of ECD.

to what households in community-based ECD villages would do if a public ECD center existed in their village. Then, the implied first-stage increase in take-up in the community-based ECD experiment, is 90% - 30% = 60%. This first stage implies that our ITT estimates should be scaled by a factor of 1.7 (1/0.6) to approximate the LATE for children induced to attend an ECD center because of treatment.

For the ECD Annex experiment, the relevant first-stage for the ECD Annex experiment is the increase in teaching inputs due to training. As discussed in Section 2, no ECD Annex site reported implementation problems, suggesting that the experiment increased input quality as intended. Furthermore, the 90% baseline ECD attendance in these communities suggests that nearly all sampled children in treated communities were exposed to treated teachers. Therefore, a good approximation to the average treatment effect of teacher training for children attending ECD Annexes is to scale the ITT by 1.1 (1/0.9).

To gain further insight into implementation, Table 4 reports data from a monitoring survey of ECD center administrators and teachers taken in 2013, midway through the program. The survey occurred only in treated communities within each experiment. The top rows of the table show that ECD centers were well attended in both experiments, with an average of more than 45 children registered per center and more than 30 children present on the day of the monitoring visit. Attendance rates on the day of the visit were significantly higher in community-based ECD centers than in treated ECD Annexes, reflecting strong demand in communities that previously lacked structured ECD services. The pupil-teacher ratio exceeds 30 in both experiments, limiting the quality of services that teachers can provide.

The bottom rows of Table 4 report teacher characteristics. Compared to treated ECD Annexes, teachers in community-based ECD centers are significantly younger, less likely to have completed secondary school, and have less teaching and ECD experience. They are also more likely to be a village resident. In other words, the modal provider in community-based ECD centers is a local volunteer without previous experience, compared to the professional staff in ECD Annexes. These differences also have implications for the expected quality of service delivery within each experiment, a subject to which we return below. Nonetheless, survey enumerators report higher levels of student engagement in community-based ECD centers than in ECD Annexes (bottom rows of table), though only the “children taking initiative” category

differs significantly. These results suggest that teacher effort in community-based ECD centers was at least as high, if not higher, than in ECD Annexes.

4.3. Endline outcomes

We report endline MDAT scores and treatment effects estimates in Table 5. The first two columns show the control and treatment mean, respectively, for the community-based ECD experiment. Columns (3) and (4) show estimates of equation (1), with and without controlling for the baseline outcome. Columns (5)-(8) present the same information for the ECD Annex experiment. In the community-based ECD experiment, the treatment group scores below the control in both language and hearing and fine motor skills. Although this pattern is surprising, neither difference is statistically distinguishable from zero, regardless of whether we control for the baseline score. Controlling for variables imbalanced at baseline shrinks estimates further towards zero, without loss of precision (see Table A1). We also fail to find significant treatment effects when disaggregating each MDAT module by subsets of items most closely associated with the ECD curriculum. Overall, we find no evidence of differences in average scores based on exposure to community-based ECD services.

When dropping sites that reported implementation issues, the treatment effect of community-based ECD on language and hearing falls to -0.24 standard deviations, significant at 10%. Including the baseline score increases the magnitude to -0.29 standard deviations, significant at 5%. The result is surprising, since we expect estimated benefits of community-based ECD to increase when omitting sites with implementation issues. A potential explanation is reporting bias: more conscientious administrators are more likely to report implementation problems, so excluding them leaves only the worst managed sites in the sample. We find no significant differences in fine motor skills when dropping sites with implementation issues. We omit results for brevity but present them in Table SA4 of the Supplemental Appendix.[17]

[17] Supplemental Appendix Table SA5 presents results when using multiple imputation to infer MDAT scores of children who did not take the test. Following Little and Rubin (2014), we use the baseline variables reported in Table 3 to predict MDAT scores of children with missing scores. We impute missing values 50 times, with inference accounting for the imputation procedure. We impute only overall scores, not subsets of test items, because we are less confident in the results at this level of granularity. Our results for overall scores are of the same signs and similar magnitudes as the main results, with the treatment effect for language in the community-based experiment negative and significant at 10%. These results suggest that the high proportion of children with missing MDAT scores reduces statistical power but does not qualitatively change results.

In the ECD Annex experiment, the treatment group scores higher on both language and hearing and fine motor skills, but the differences are not precisely estimated. Among subsets of items, the only significant differences are for counting and ordering rows of items, though here the treatment group scores below the control, by 9 and 8 percentage points, respectively.

The mean effects presented in Table 5 might mask changes in other features of the outcome distribution between children in the treatment and control groups of each experiment. We explore this possibility in Figure 4, which plots the distributions of MDAT scores. Figure 4(a) shows kernel density estimates for the community-based ECD experiment, with baseline scores plotted in the first row and endline scores in the second row. Figure 4(b) shows the corresponding density estimates for the ECD Annex experiment. Comparing densities within a column shows how the distributions of each MDAT module shift between baseline and endline.[18]

Beginning with fine motor skills for the community-based ECD experiment in the first column of Figure 4(a), the treatment group lies slightly to the left of the control distribution at baseline. At endline, the mode of the treatment group distribution is to the right of the control group mode. The treatment group's thicker left tail and thinner right tail make the overall change unclear, however. For language skills in the second column, again the treatment group distribution lies slightly to the left of the control group, with only minor differences apparent at endline. In short, we fail to find strong evidence of relative shifts in the MDAT score distributions in the community-based ECD experiment.

Turning to the ECD Annex experiment in Figure 4(b), at baseline we see similar central locations for fine motor skills between the treatment and control groups, though the treatment group has a thicker right tail. By the endline, the treatment group distribution has shifted right relative to the control. In language and hearing, the treatment group begins to the right of the control group distribution at baseline, with the difference somewhat more pronounced at endline. These differences are consistent with the positive point estimates for the ECD Annex treatment coefficient in Table 5.

Although suggestive, the preceding visual inspection of unconditional distributions should not substitute for formal analysis. Even formal tests for equality of unconditional

[18] Because we rely on different versions of the MDAT in baseline and endline, the densities are informative about relative changes in the treatment and control distributions, but not of absolute changes in child development.

distributions would fail to account for the stratification of treatment assignment by region, or for differences in baseline outcomes. To overcome these limitations, we estimate equation (1) using quantile regressions. For each MDAT module, we estimate the coefficient on treatment status for the 5^{th}-95^{th} (conditional) quantiles, in increments of 5. We continue to cluster standard errors by community, the unit of treatment.

Figure 5(a) presents results for the community-based ECD experiment, with results for the ECD Annex experiment in Figure 5(b). In each graph, the thick black line plots our estimates of the quantile treatment effects. The gray shaded area shows the 95% confidence interval around these estimates. For comparison, we also plot the OLS estimate of the average treatment effect in the long-dashed line, while the short-dashed line shows its 95% confidence interval. The red line shows the gridline for zero treatment effect.

For the community-based ECD experiment in Figure 5(a), a downward slope appears in quantile treatment effect estimates of fine motor skills beginning around the 40^{th} percentile. For language and hearing, estimates remain near the OLS coefficient across all quantiles. For the ECD Annex experiment in Figure 5(b), a downward slope appears in quantile treatment effect estimates around the 40^{th} percentile for fine motor skills and the 20^{th} percentile for language and hearing. These downward slopes suggest that treatment benefitted the middle of the outcome distributions most, with smaller benefits at the top of the distribution. Across both experiments and outcomes, however, estimates of quantile treatment effects are noisy. In fact, confidence intervals include zero around all the quantile effects estimated, suggesting that any differences between treatment and control distributions implied by Figure 4 are spurious. Results are similar when augmenting the quantile regression specification with baseline outcomes, or when using the nonparametric quantile treatment effect estimator of Firpo (2007); see Figure SA1 of the Supplemental Appendix.

4.3 Heterogeneous treatment effects

Although we have found no evidence of significant treatment effects for our main measures of childhood development in the analysis thus far, the null effects for each experimental sample might mask significant effects for subgroups of children. This analysis is more exploratory since, although we intended to explore heterogeneous impacts according to child and household characteristics, our study was not explicitly designed to detect them. We

examine the following subgroups, defined according to baseline characteristics: male and female; below and above median household assets; whether the child's mother ever attended school; below or above median quantity of stimulating objects in the home (based on a predetermined list of items); whether the child was sick in the past three days; and below and above median mother mental distress, created from a series of questions about mental health. We chose these groups to analyze whether treatment effects vary by child sex, socioeconomic status, and child and parental health. Subgroups defined as below/above median use both experimental samples to determine the threshold; results are similar when defining the median within each experimental sample.

For each pair of subgroups, we estimate the regression:

$$y_{ic} = \beta_a D_c * \mathbf{1}(\text{group } a)_{ic} + \beta_b D_c * \mathbf{1}(\text{group } b)_{ic} + X_{ic}\theta + \gamma Region2_c + \varepsilon_{ic} \quad (2)$$

where $D_c * \mathbf{1}(\text{group } a)_{ic}$ is the interaction between treatment and membership in group *a* (e.g., male), $D_c * \mathbf{1}(\text{group } b)_{ic}$ is the analogous term for group *b* (e.g., female), and *X* is a vector of membership dummies for *all* subgroups listed above (not merely those included in the interactions). In other words, we run several versions of equation (2) in order to test for different subgroup interactions, but all control variables remain the same across equations. Our coefficients of interest are β_a and β_b, which measure differential treatment effects for groups *a* and *b* relative to those same groups of children in the control group. For instance, when groups *a* and *b* are male and female, then β_a and β_b measure treatment effects for male and female children, respectively (the constant is omitted for identification and ease of interpretation).

Within each experiment and child development outcome (fine motor skills and language and hearing), we stack equation (2) across all subgroups and estimate the system jointly. This specification accounts for correlations among error terms in the system and permits hypothesis testing across equations. In addition to testing for significance of all subgroup interaction terms, we are interested in how treatment effects vary by household socioeconomic status (SES). We define low SES as below median household assets, mother did not attend school, and below median stimulating objects in the home, while high SES is the complement of these groups. We then conduct the following joint hypothesis tests:

1. *H1:* $\beta_{low\ assets} = \beta_{mother\ no\ school} = \beta_{low\ stimulation} = 0$

2. *H2:* $\beta_{high\ assets} = \beta_{mother\ attended\ school} = \beta_{high\ stimulation} = 0$
3. *H3:* $(\beta_{low\ assets} + \beta_{mother\ no\ school} + \beta_{low\ stimulation}) =$
   $(\beta_{high\ assets} + \beta_{mother\ attended\ school} + \beta_{high\ stimulation})$

Hypothesis *H1* tests whether children from low-SES households have non-zero treatment effects. Hypothesis *H2* is the analogous test for children from high-SES households. Hypothesis *H3* tests whether treatment effects for low- and high-SES groups differ from each other.

Table 6(a) presents estimates of equation (2) for the community-based ECD experiment, with p-values of hypotheses *H1-H3* reported at the bottom of the table. We find negative and statistically significant treatment effects for several subgroups: in fine motor skills, for those with above-median stimulating objects at home; and in language, for females, above-median household assets, mother attended school, above-median stimulating objects at home, child not sick in last three days, and both below- and above-median mother mental distress. The magnitudes range from -0.2 to -0.4 standard deviations. The pattern suggests that females and children from more advantaged households experienced slower language and hearing development via exposure to the community-based ECD program. Our joint hypothesis tests confirm statistically significant negative treatment effects in language for high-SES children. We also find that high-SES children have smaller treatment effects than low-SES children in both fine motor skills and language. Although surprising, these results are plausible if better-off households are able to provide a higher quality home environment than the community-based alternative. A related possibility is that the curriculum was not sufficiently targeted to improving the skills measured by the MDAT, resulting in slower development for children from homes better able to foster these particular skills.

For the ECD Annex experiment (Table 6(b)), most point estimates are positive, but few subgroups of children have treatment effects significantly different from zero. None of the joint hypothesis tests is statistically significant. We conclude that there is no pattern of treatment effect heterogeneity in the ECD Annex experiment among the groups considered.

A potential concern with these results is that pre-existing differences at baseline between treatment and control has led to spurious impact estimates. Of particular concern is the relatively large discrepancy in baseline ECD attendance between groups in the community-based ECD experiment, documented in Table 3. To explore this possibility, we added baseline ECD

attendance and non-attendance as an additional subgroup pair when estimating equation (2). Additionally, we include all variables imbalanced at baseline in either experiment as additional controls. We present results in Table A2. Interactions between baseline ECD attendance and treatment are not statistically significant in either experiment. Moreover, the pattern of differential results by baseline SES remains unchanged for the community-based ECD experiment, though somewhat less precise. The p-value on the joint hypothesis test that children from high-SES households have a non-zero treatment effect for language in the community-based ECD experiment is 0.14, though differences between high- and low-SES children remain significant (p=.03). We conclude that baseline imbalances between treatment and control children do not invalidate our main results.

4.4. Home investments

The final outcome we consider is home investments. Parents of children exposed to either treatment may alter their investment in their children in response. Specifically, the focus of the new GOALS curriculum on stimulation through structured play may encourage parents to increase play activities at home. We therefore explore whether treatment assignment increases the time mothers report playing with their children. It is also possible that parents substitute away from private investments in children in response to higher public investments, in the form of ECD centers. Such a mechanism could even be strong enough generate negative total impacts of these interventions on child outcomes.

We use the full sample of children, regardless of whether they completed the MDAT, in order to maximize statistical power.[19] Table 7, column (1) presents estimates of equation (1), using mother's play time (minutes/day) as the outcome and assignment to community-based ECD as the treatment. The coefficient on treatment is positive, but small (5 minutes, compared to the control mean of 184) and not statistically significant. Columns (2)-(7) present estimates of equation (2), following the same format as Table 6. None of the subgroup treatment effect estimates is significantly different from zero, either individually or jointly. These results suggest

[19] Supplemental Appendix Table SA6 presents robustness checks using the MDAT sample and an additional measure of home investment (mother's total interaction time with children). Results are similar across these alternative specifications.

that negative treatment effects for some subgroups exposed to community-based ECD are not driven by reductions in home investments.

Columns (8)-(14) of Table 7 repeat the exercise for the ECD Annex experiment. Although the overall treatment effect is not statistically distinguishable from zero, we find significant increases in play time for several subgroups. These effects are large, ranging from 36-104 minutes, relative to the control group mean of 162. These increases in home investments make the null findings for the ECD Annex treatment more notable, as the combination of increased teacher training and home-based stimulation failed to improve measures of childhood development.

Table 8, column (1) presents estimates of equation (1), using stimulating objects (measured as the proportion from a predetermined list) as the outcome and assignment to community-based ECD as the treatment. The coefficient on treatment is positive, but small (3 percent points, compared to the control mean of 30 percent) and not statistically significant. Columns (2)-(7) present estimates of equation (2), again following the same format as Table 6. Although treatment effects are positive and significant for male children and households with below-median assets, these are not the groups for whom we found significant treatment effects in Table 6. These results suggest that negative treatment effects for some subgroups exposed to community-based ECD are not driven by reductions in home investments.

Columns (8)-(14) of Table 8 repeat the exercise for the ECD Annex experiment. We find no significant treatment effects for any group, either individually or jointly. These results are consistent with our null findings for the effect of this intervention.

4.5. Discussion

Tables 7-8 show that exposure to the community-based ECD treatment did not reduce their household investments in children. We are left with a puzzle: why were there negative effects of community-based ECD exposure among relatively advantaged households? A potential explanation is that these households were steered away from better quality early childhood settings for their children, possibly in their homes. To explore this possibility, we split the sample by whether the child attended ECD at baseline and re-estimate the treatment effect heterogeneity specifications of Table 6. If the counterfactual home environment is driving results, we should see negative effects of the community-based ECD treatment among the

subsample whose children were home at baseline. Indeed, this is what we find, with significant and negative coefficients among advantaged households whose children were home at baseline (Table A3(a)), but no significant effects among children already attending ECD at baseline (Table A3(b)). We concede, however, that point estimates across the two groups are similar, and p-values of the joint tests by socioeconomic status are also inconclusive. Nonetheless, the pattern of results is consistent with our interpretation.

If community-based ECD centers were of such low quality, why would parents send their children? Perhaps they held an erroneous perception of classroom quality, given that the facilities were new. Another possibility is that they valued the time freed up by sending their children to preschool instead of caring for them at home. We fail to find increases in household labor force attachment or decreases in mental distress in response to the treatment, though utility of households' existing time allocation may have increased in other ways.[20] We lack data on household perceptions of ECD quality to distinguish between these explanations.

5. Conclusion

This paper evaluated two experiments to improve early childhood development services in The Gambia. The first experiment focused on increasing access to ECD services by constructing community-based centers in communities where no structured ECD program existed. The second experiment focused on improving quality of existing ECD centers by training teachers to deliver a new curriculum. We found no evidence that either intervention affected average levels of child development, as measured by tests of fine motor skills or language and hearing. Although the experimental evidence comes from only two of six regions in the country, the same programs were designed by the government to be scaled up nationally, with the annexes covering more regions at the time of this study. The scale of these programs make these findings directly relevant for policy.

Exploratory analysis suggests that children from more advantaged households developed language skills more slowly when exposed to the community-based ECD treatment. Parents of

[20] Results in supplemental appendix Table SA7. Consistent with this interpretation, household expenditure rose 65 USD in response to treatment, suggesting the productivity of existing work increased. However, the result is significant only at 10% and becomes insignificant when controlling for household expenditure at baseline.

these children did not change their home investments, suggesting that these effects are driven by the low quality of community-based ECD centers, not by household responses to treatment.

Our results align with other studies finding modest, or even negative, effects of center-based ECD programs when implemented at scale. Baker, Gruber, and Milligan (2008, 2015) find that the introduction of universal child care in Canada led to negative effects on non-cognitive skills, with consequences persisting into adulthood. Ichino, Fort, and Zanella (2019) also find negative effects of center-based daycare in Italy, with the largest negative effects for girls and more affluent households. In developing countries, Rosero and Oosterbeek (2011) and Bernal et al. (2019) find negative effects of center-based care in Ecuador and Colombia, respectively, as do Bouguen et al. (2018) for preschool construction in Cambodia. Unifying these studies and ours is that the programs are publicly provided and large scale. Such scale creates challenges for governments to maintain quality relative to care at home, particularly for relatively better-off families or in poor countries. To be effective, scaled programs in poor countries might need to include nutrition or health components in addition to education, as in the program in Colombia studied in Bernal and Ramírez (2019). Taken together, these results should sound a note of caution for governments seeking to implement center-based care at large scale.

Moreover, given the importance of early childhood circumstances for adult outcomes, longer-term measures are required to account fully for the effects of the interventions studied here. We are only able to document short-term impact estimates, focusing on young children less than two years after treatment assignment.

References

Abadie, Alberto. 2005. “Semiparametric Difference-in-Differences Estimators.” *The Review of Economic Studies* 72 (1): 1–19.

Aboud, Frances E. 2006. “Evaluation of an Early Childhood Preschool Program in Rural Bangladesh.” *Early Childhood Research Quarterly* 21 (1): 46–60.

Armecin, Graeme, Jere Behrman, Paulita Duazo, Sharon Ghuman, Socorro Gultiano, Elizabeth King, and Nannette Lee. 2006. “Early Childhood Development through an Integrated Program: Evidence from the Philippines.” *World Bank Policy Research Working Paper*, no. 3922. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=917493.

Baker, Michael, Jonathan Gruber, and Kevin Milligan. 2008. “Universal Child Care, Maternal Labor Supply, and Family Well-Being.” *Journal of Political Economy* 116 (4): 709–745.

———. 2015. “Non-Cognitive Deficits and Young Adult Outcomes: The Long-Run Impacts of a Universal Child Care Program.” Working Paper 21571. National Bureau of Economic Research. https://doi.org/10.3386/w21571.

Baker-Henningham, Helen, Stephen Scott, Kelvyn Jones, and Susan Walker. 2012. “Reducing Child Conduct Problems and Promoting Social Skills in a Middle-Income Country: Cluster Randomised Controlled Trial.” *The British Journal of Psychiatry* 201 (2): 101–108.

Bastos, Paulo, Nicolas L. Bottan, and Julian Cristia. 2017. “Access to Preprimary Education and Progression in Primary School: Evidence from Rural Guatemala.” *Economic Development and Cultural Change*, February, 000–000. https://doi.org/10.1086/691090.

Behrman, Jere, Lia Fernald, and Patrice Engle. 2013. “Preschool Programs in Developing Countries.” *Education Policy in Developing Countries*, 65–105.

Behrman, Jere R., Yingmei Cheng, and Petra E. Todd. 2004. “Evaluating Preschool Programs When Length of Exposure to the Program Varies: A Nonparametric Approach.” *Review of Economics and Statistics* 86 (1): 108–132.

Berlinski, S., and N. Schady. 2015. *The Early Years: Child Well-Being and the Role of Public Policy. Development in the Americas Series*. New York: Palgrave MacMillan, and Washington, DC: Inter-American Development Bank.

Berlinski, Samuel, and Sebastian Galiani. 2007. “The Effect of a Large Expansion of Pre-Primary School Facilities on Preschool Attendance and Maternal Employment.” *Labour Economics* 14 (3): 665–680.

Berlinski, Samuel, Sebastian Galiani, and Paul Gertler. 2009. “The Effect of Pre-Primary Education on Primary School Performance.” *Journal of Public Economics* 93 (1–2): 219–34. http://www.sciencedirect.com/science/journal/00472727.

Berlinski, Samuel, Sebastian Galiani, and Marco Manacorda. 2008. “Giving Children a Better Start: Preschool Attendance and School-Age Profiles.” *Journal of Public Economics* 92 (5–6): 1416–40. http://www.sciencedirect.com/science/journal/00472727.

Bernal, Raquel. 2015. “The Impact of a Vocational Education Program for Childcare Providers on Children’s Well-Being.” *Economics of Education Review* 48 (October): 165–83. http://www.sciencedirect.com/science/journal/02727757.

Bernal, Raquel, Orazio Attanasio, Ximena Peña, and Marcos Vera-Hernández. 2019. “The Effects of the Transition from Home-Based Childcare to Childcare Centers on Children’s Health and Development in Colombia.” *Early Childhood Research Quarterly* 47: 418–431.

Bernal, Raquel, and Camila Fernández. 2013. “Subsidized Childcare and Child Development in Colombia: Effects of Hogares Comunitarios de Bienestar as a Function of Timing and Length of Exposure.” *Social Science & Medicine* 97: 241–249.

Bernal, Raquel, and Sara María Ramírez. 2019. “Improving the Quality of Early Childhood Care at Scale: The Effects of ‘From Zero to Forever.’” *World Development* 118 (June): 91–105. https://doi.org/10.1016/j.worlddev.2019.02.012.

Black, Maureen M, Susan P Walker, Lia C H Fernald, Christopher T Andersen, Ann M DiGirolamo, Chunling Lu, Dana C McCoy, et al. 2017. “Early Childhood Development Coming of Age: Science through the Life Course.” *The Lancet* 389 (10064): 77–90. https://doi.org/10.1016/S0140-6736(16)31389-7.

Blimpo, Moussa, Pedro Carneiro, and Nathalie Lahire. 2016. “Improving Parental Investments in Children: Experimental Evidence from The Gambia,” March.

Blimpo, Moussa P., Pedro Carneiro, Pamela Jervis Ortiz, and Todd Pugatch. 2018. “Improving Parental Investments in Children: Experimental Evidence from The Gambia.”

Bouguen, Adrien, Deon Filmer, Karen Macours, and Sophie Naudeau. 2014. “Preschools and Early Childhood Development in a Second Best World: Evidence from a Scaled-up Experiment in Cambodia.” http://www.cepr.org/active/publications/discussion_papers/dp.php?dpno=10170.

———. 2018. “Preschool and Parental Response in a Second Best World Evidence from a School Construction Experiment.” *Journal of Human Resources* 53 (2): 474–512. https://doi.org/10.3368/jhr.53.2.1215-7581R1.

Bowne, Jocelyn Bonnes, Hirokazu Yoshikawa, and Catherine E. Snow. 2016. “Experimental Impacts of a Teacher Professional Development Program in Early Childhood on Explicit Vocabulary Instruction across the Curriculum.” *Early Childhood Research Quarterly* 34: 27–39.

Currie, Janet, and Douglas Almond. 2011. “Human Capital Development before Age Five.” *Handbook of Labor Economics* 4: 1315–1486.

Firpo, Sergio. 2007. “Efficient Semiparametric Estimation of Quantile Treatment Effects.” *Econometrica* 75 (1): 259–276.

Gambia Ministry of Basic and Secondary Education. 2017. “Education Sector Strategic Plan 2016 – 2030.” https://www.globalpartnership.org/sites/default/files/2018-09-the-gambia-essp-2016-30.pdf.

Giordono, Leanne, and Todd Pugatch. 2017. “Non-Tuition Costs, School Access and Student Performance: Evidence from the Gambia.” *Journal of African Economies*. http://jae.oxfordjournals.org/content/early/2017/01/05/jae.ejw033.abstract.

Gladstone, Melissa, Gillian A. Lancaster, Eric Umar, Maggie Nyirenda, Edith Kayira, Nynke R. van den Broek, and Rosalind L. Smyth. 2010. “The Malawi Developmental Assessment Tool (MDAT): The Creation, Validation, and Reliability of a Tool to Assess Child Development in Rural African Settings.” *PLOS Med* 7 (5): e1000273. https://doi.org/10.1371/journal.pmed.1000273.

He, Fang, Leigh L. Linden, and Margaret MacLeod. 2009. “A Better Way to Teach Children to Read? Evidence from a Randomized Controlled Trial.” *Unpublished Manuscript. New York, NY: Columbia University*. http://leighlinden.com/Teach%20Children%20to%20Read.pdf.

Hirano, Keisuke, Guido W. Imbens, and Geert Ridder. 2003. “Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score.” *Econometrica* 71 (4): 1161–1189.

Ichino, Andrea, Margherita Fort, and Giulio Zanella. 2019. “Cognitive and Non-Cognitive Costs of Daycare 0-2 for Children in Advantaged Families.” *Journal of Political Economy*, April. https://doi.org/10.1086/704075.

Imbens, Guido W., and Jeffrey M. Wooldridge. 2009. “Recent Developments in the Econometrics of Program Evaluation.” *Journal of Economic Literature* 47 (1): 5–86. http://www.aeaweb.org/jel/index.php.

Krafft, Caroline. 2015. “Increasing Educational Attainment in Egypt: The Impact of Early Childhood Care and Education.” *Economics of Education Review* 46: 127–143.

Lee, David S. 2009. “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects.” *The Review of Economic Studies* 76 (3): 1071–1102.

Little, Roderick JA, and Donald B. Rubin. 2014. *Statistical Analysis with Missing Data*. Vol. 333. John Wiley & Sons.

Malmberg, Lars-Erik, Peter Mwaura, and Kathy Sylva. 2011. “Effects of a Preschool Intervention on Cognitive Development among East-African Preschool Children: A Flexibly Time-Coded Growth Model.” *Early Childhood Research Quarterly* 26 (1): 124–33. https://doi.org/10.1016/j.ecresq.2010.04.003.

Martinez, Sebastian, Sophie Naudeau, and Vitor Pereira. 2013. “The Promise of Preschool in Africa: A Randomized Impact Evaluation of Early Childhood Development in Rural Mozambique,” November. http://www.fucape.br/downloads/Artigo%20Vitor%20Azevedo.pdf.

Mwaura, Peter A.M., Kathy Sylva, and Lars-Erik Malmberg. 2008. “Evaluating the Madrasa Preschool Programme in East Africa: A Quasi-experimental Study.” *International Journal of Early Years Education* 16 (3): 237–55. https://doi.org/10.1080/09669760802357121.

Ozler, Berk, Lia C. H. Fernald, Patricia Karol Kariger, Christin Mcconnell, Michelle J. Neuman, and Eduardo Fraga. 2016. “Combining Preschool Teacher Training with Parenting Education: A Cluster-Randomized Controlled Trial.” SSRN Scholarly Paper ID 2849117. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=2849117.

Pinto, Cristine Campos de Xavier, Daniel Santos, and Clarissa Guimarães. 2016. “The Impact of Daycare Attendance on Math Test Scores for a Cohort of Fourth Graders in Brazil.” *The Journal of Development Studies*, 1–23.

Raine, Adrian, Jianghong Liu, Peter H. Venables, Sarnoff A. Mednick, and C. Dalais. 2010. “Cohort Profile: The Mauritius Child Health Project.” *International Journal of Epidemiology* 39 (6): 1441–1451.

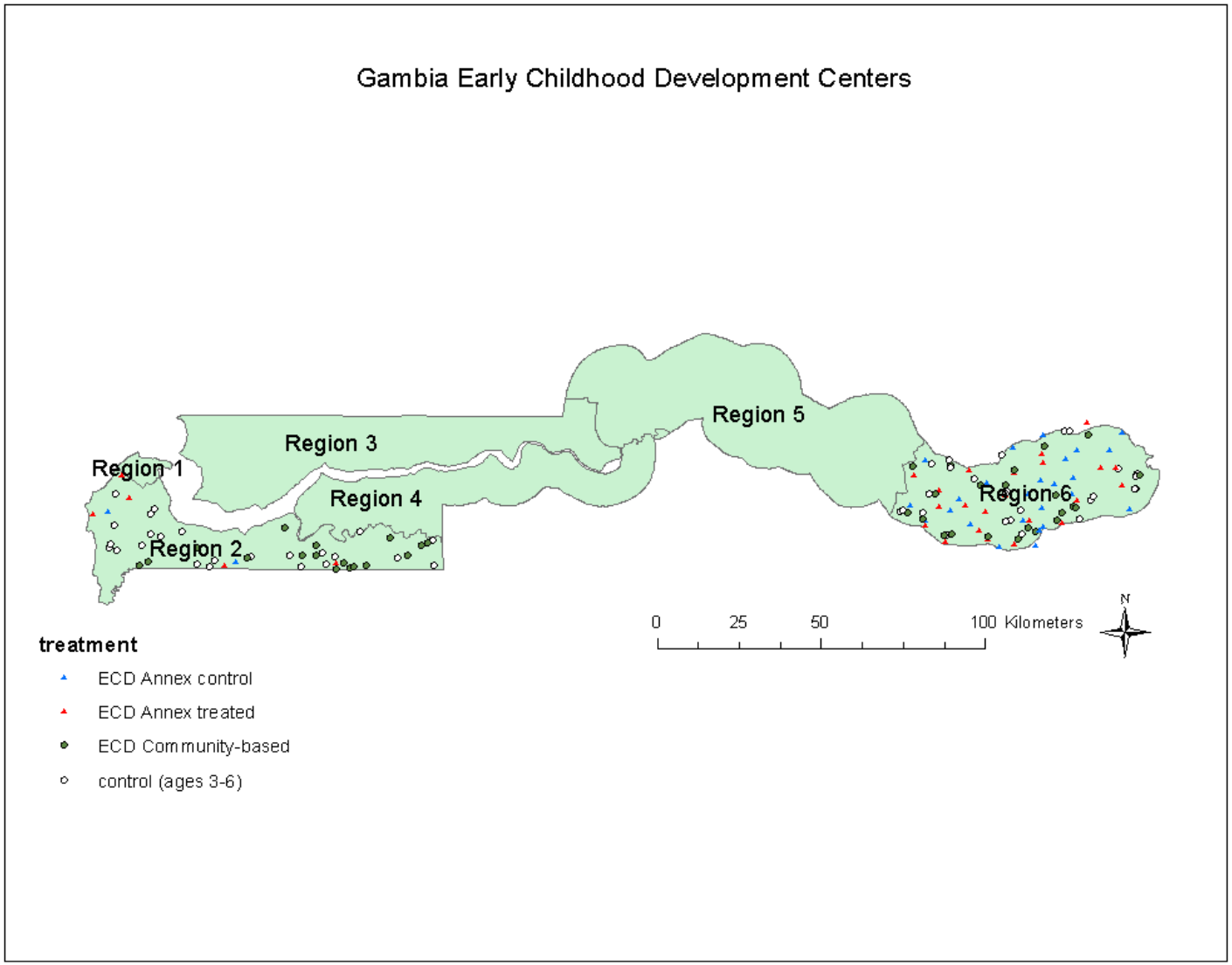
Raine, Adrian, Kjetil Mellingen, Jianghong Liu, Peter Venables, and Sarnoff A. Mednick. 2003. “Effects of Environmental Enrichment at Ages 3-5 Years on Schizotypal Personality and Antisocial Behavior at Ages 17 and 23 Years.” *American Journal of Psychiatry* 160 (9): 1627–1635.

Raine, Adrian, Peter H. Venables, Cyril Dalais, Kjetil Mellingen, Chandra Reynolds, and Sarnoff A. Mednick. 2001. “Early Educational and Health Enrichment at Age 3–5 Years Is Associated with Increased Autonomic and Central Nervous System Arousal and

Orienting at Age 11 Years: Evidence from the Mauritius Child Health Project." *Psychophysiology* 38 (02): 254–266.
Raine, Adrian, Peter H. Venables, and Sarnoff A. Mednick. 1997. "Low Resting Heart Rate at Age 3 Years Predisposes to Aggression at Age 11 Years: Evidence from the Mauritius Child Health Project." *Journal of the American Academy of Child & Adolescent Psychiatry* 36 (10): 1457–1464.
Rao, Nirmala, Jin Sun, Veronica Pearson, Emma Pearson, Hongyun Liu, Mark A. Constas, and Patrice L. Engle. 2012. "Is Something Better than Nothing? An Evaluation of Early Childhood Programs in Cambodia." *Child Development* 83 (3): 864–876.
Robins, James M., and Andrea Rotnitzky. 1995. "Semiparametric Efficiency in Multivariate Regression Models with Missing Data." *Journal of the American Statistical Association* 90 (429): 122–129.
Rosero, Jose. 2012. "On the Effectiveness of Child Care Centers in Promoting Child Development in Ecuador." Tinbergen Institute. http://papers.tinbergen.nl/12075.pdf.
Rosero, José, and Hessel Oosterbeek. 2011. "Trade-Offs between Different Early Childhood Interventions: Evidence from Ecuador." http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1898566.
Taiwo, A. A., and J. B. Tyolo. 2002. "The Effect of Pre-School Education on Academic Performance in Primary School: A Case Study of Grade One Pupils in Botswana." *International Journal of Educational Development* 22 (2): 169–180.
Tanner, Jeffrey C., Tara Candland, and Whitney S. Odden. 2015. "Later Impacts of Early Childhood Interventions: A Systematic Review," World Bank Independent Evaluation Group Working Paper, , no. 3. https://ieg.worldbankgroup.org/Data/reports/ImpactsofInterventions_EarlyChildhoodonLaterOutcomes.pdf.
Walker, Susan P., Theodore D. Wachs, Julie Meeks Gardner, Betsy Lozoff, Gail A. Wasserman, Ernesto Pollitt, Julie A. Carter, and International Child Development Steering Group. 2007. "Child Development: Risk Factors for Adverse Outcomes in Developing Countries." *The Lancet* 369 (9556): 145–157.
Walker, Susan P., Theodore D. Wachs, Sally Grantham-McGregor, Maureen M. Black, Charles A. Nelson, Sandra L. Huffman, Helen Baker-Henningham, et al. 2011. "Inequality in Early Childhood: Risk and Protective Factors for Early Child Development." *The Lancet* 378 (9799): 1325–1338.
Wolf, Sharon, J. Lawrence Aber, and Jere R. Behrman. 2017. "The Impacts of Teacher Training and Parental Education on Kindergarten Quality in Ghana." Innovations for Poverty Action Study Summary. https://www.poverty-action.org/printpdf/20016.
World Bank. 2015. "World Development Indicators." World Bank. http://data.worldbank.org/data-catalog/world-development-indicators.
———. 2016. "The Gambia: Country at a Glance." http://www.worldbank.org/en/country/gambia.
Yoshikawa, Hirokazu, Diana Leyva, Catherine E. Snow, Ernesto Treviño, M. Barata, Christina Weiland, Celia J. Gomez, et al. 2015. "Experimental Impacts of a Teacher Professional Development Program in Chile on Preschool Classroom Quality and Child Outcomes." *Developmental Psychology* 51 (3): 309.
Zoyem, Jean-Paul. 2010. "Design and Support to the Impact Evaluation for the Early Childhood Care Development (ECCD) in Gambia: Inception Report."

Zuilkowski, Stephanie Simmons, Gunther Fink, Corrina Moucheraud, and Beatrice Matafwali. 2012. “Early Childhood Education, Child Development and School Readiness: Evidence from Zambia.” *South African Journal of Childhood Education* 2 (2): 20.

Figure 1: Sample locations



Source: Authors' elaboration using school location data from Gambia Ministry of Basic and Secondary Education (MoBSE).

Figure 2: Project and research timeline

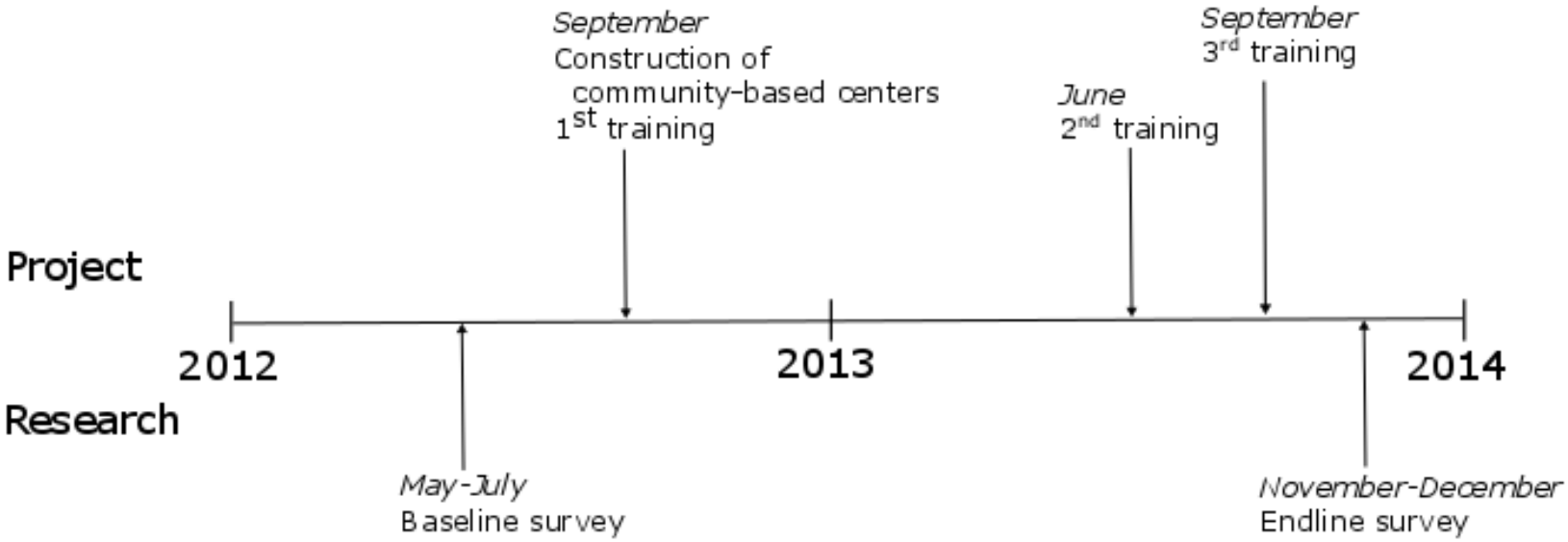


Figure 3(a): Baseline MDAT scores, by age

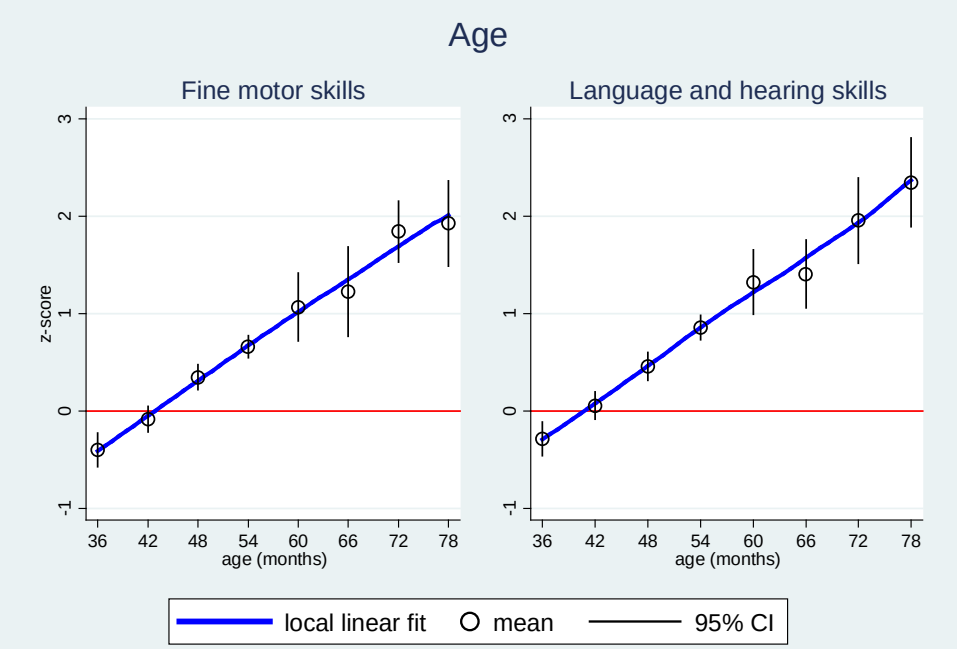
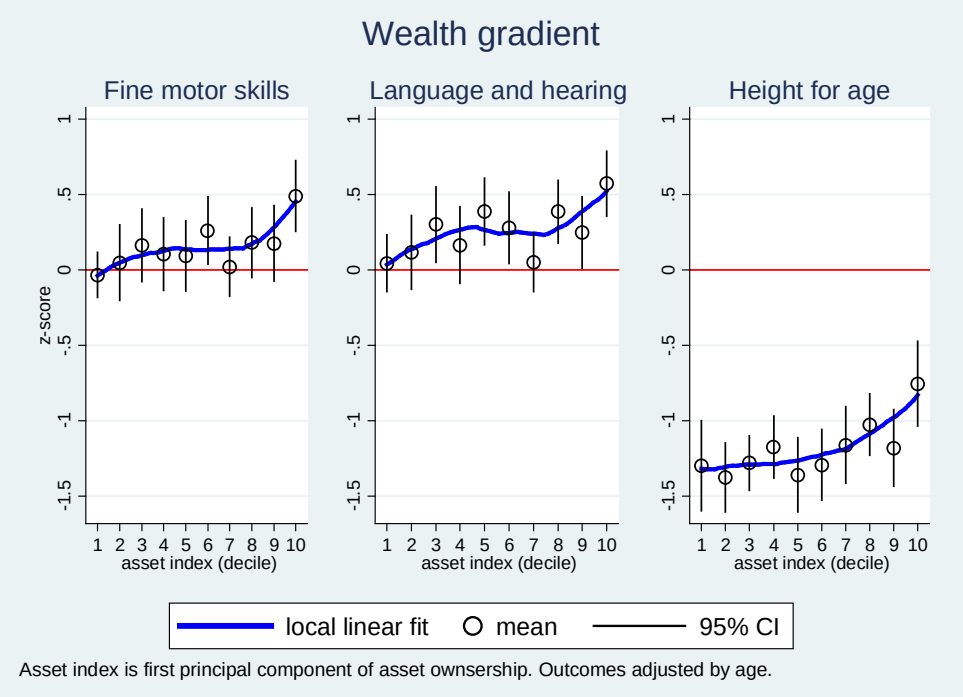


Figure 3(b): Baseline MDAT scores and height-for-age, by wealth



Notes: Figure 3(a) plots unadjusted MDAT scores by child age. Figure 3(b) plots adjusted MDAT scores and height-for-age against deciles of a household wealth index. The wealth index is the first principal component of household asset ownership. Figure 3(b) adjusts MDAT scores for age by regressing a child's raw score on age (in months), age squared, and a female dummy, then dividing the residual by its standard deviation. We run this procedure separately for the fine motor and language and hearing modules. Height-for-age calculated using the international benchmarks defined by the World Health Organization.

Figure 4(a): Outcome distributions, Community-based ECD experiment

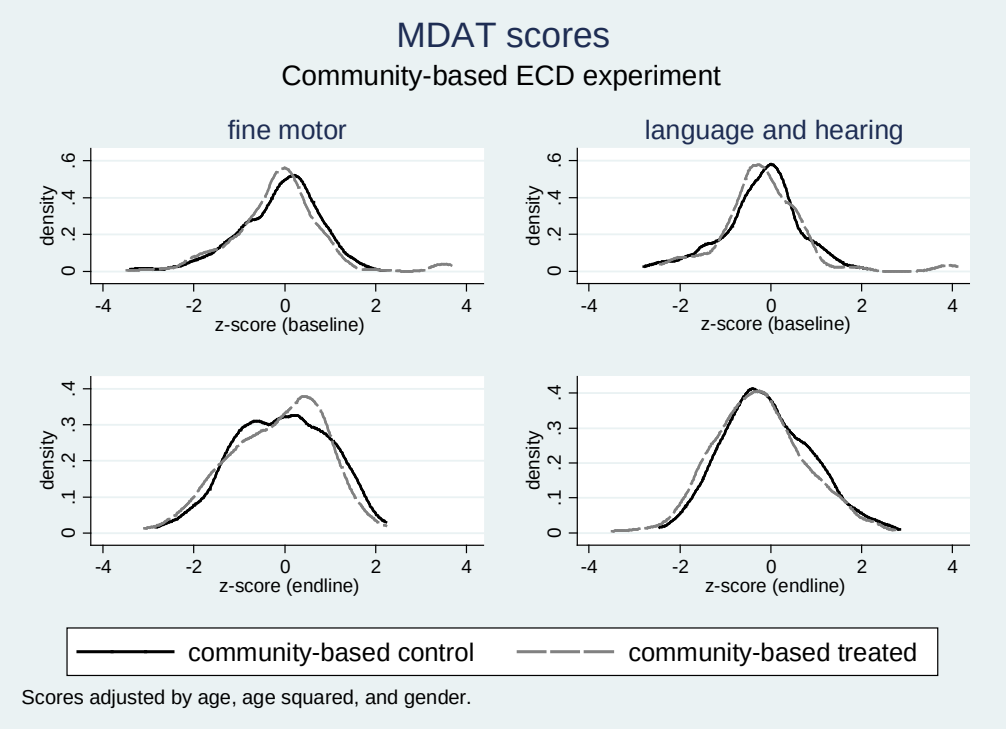
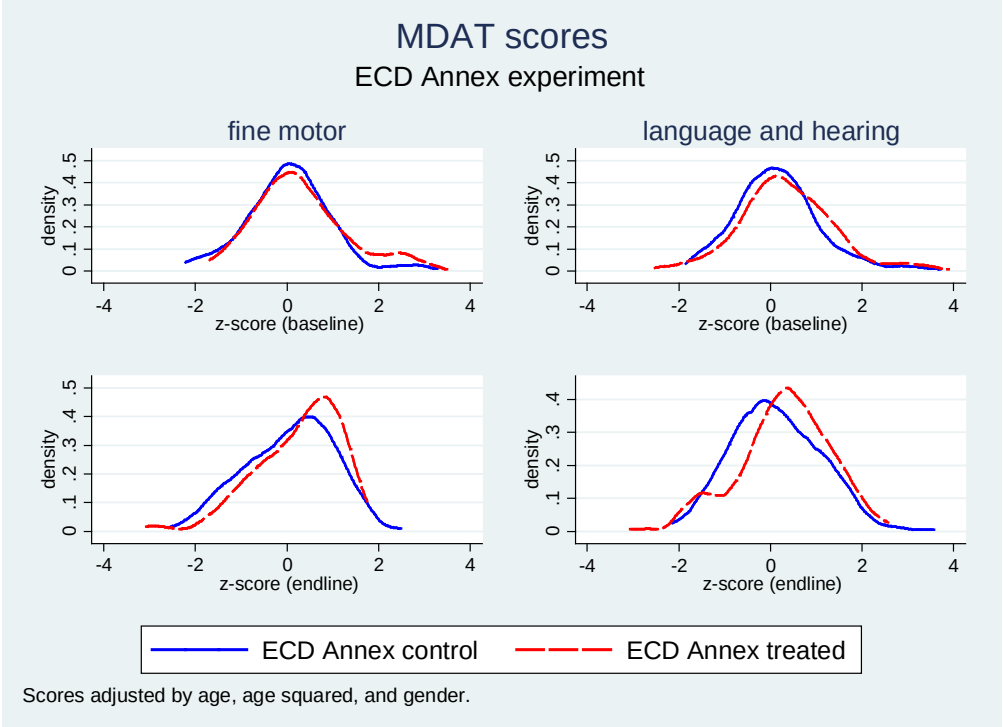


Figure 4(b): Outcome distributions, ECD Annex experiment



Notes: Figure shows kernel density estimates of MDAT scores by survey wave and experimental subgroup. MDAT scores adjusted for age and gender by regressing a child's raw score on age (in months), age squared, and a female dummy, then dividing the residual by its standard deviation. We run this procedure separately for the fine motor and language and hearing modules and for baseline and endline. Density estimates use Epanechnikov kernel with plug-in bandwidth.

Figure 5(a): Quantile treatment effects, Community-based ECD experiment

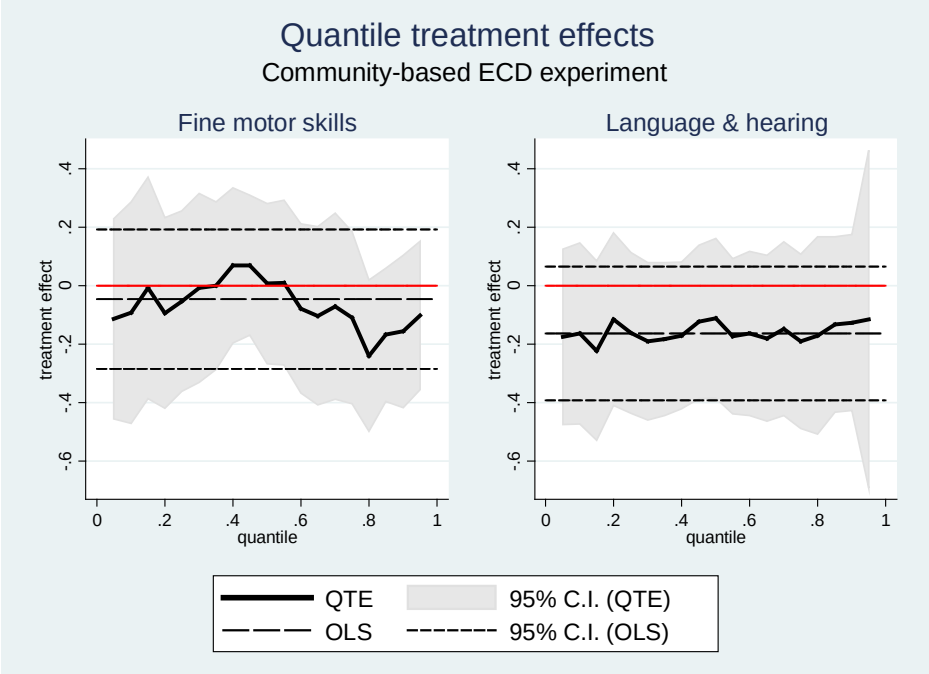
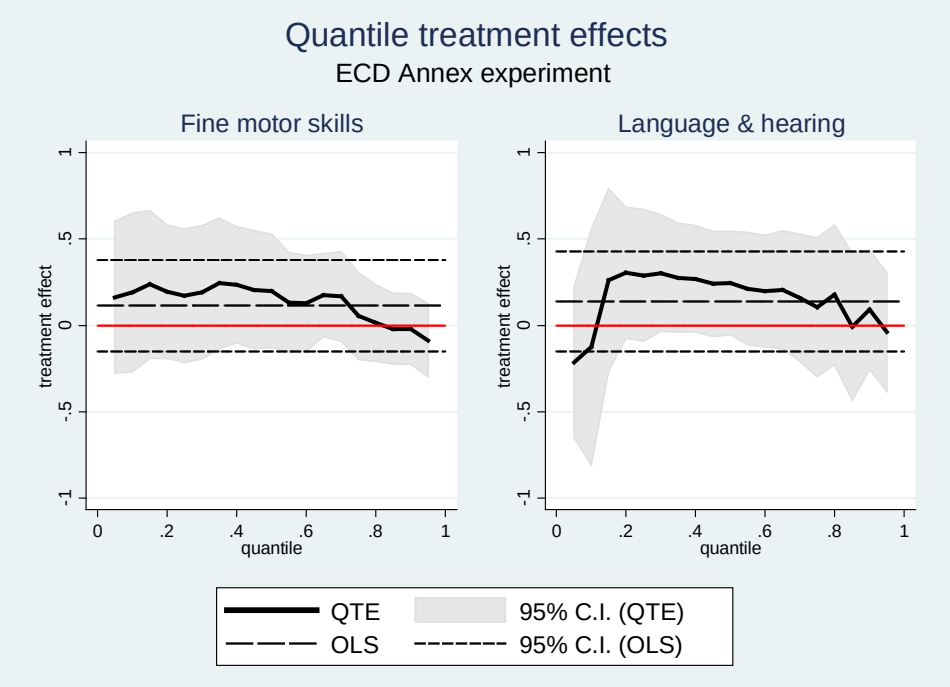


Figure 5(b): Quantile treatment effects, ECD Annex experiment



Notes: Graphs plot the coefficient on treatment status in quantile regressions of equation (1) for the 5th-95th (conditional) quantiles, in increments of 5. Standard errors clustered by community. The thick black line plots estimates of the quantile treatment effects, with 95% confidence interval shaded gray. Graphs also plot the OLS estimate in the long-dashed line (95% confidence interval in short-dashed line). The red line shows the gridline for zero treatment effect.

Table 1: Experimental design

Experiment	Sample	Control	Treatment
1	Villages without pre-existing ECD services, Regions 2/6	Pure control • No structured ECD services	Community-based • Construction of ECD center • GOALS curriculum • Teacher training • Management committee
2	Villages with ECD Annex centers, Regions 2/6	ECD Annex control • GOALS curriculum	ECD Annex treatment • GOALS curriculum • Teacher training

Table 2: Sample sizes and attrition

	(1)	(2)	(3)	(4)	(5)
Panel A: sample sizes	interview		test score		
	baseline	endline	baseline	Endline	Both
Community-based ECD experiment					
control					
original sample	606	481	319	270	243
added sample		133		75	
treatment					
original sample	441	356	267	226	204
added sample		87		55	
ECD Annex experiment					
control					
original sample	365	322	192	183	170
added sample		44		24	
treatment					
original sample	351	326	182	165	162
added sample		40		23	
Panel B: attrition	control	treatment	difference	p-value	
Community-based ECD experiment					
from endline survey	0.21	0.19	-0.01	0.99	
	(0.03)	(0.04)	(0.05)		
from endline test	0.24	0.24	0.00	0.82	
	(0.04)	(0.04)	(0.06)		
number of communities	51	40			
ECD Annex experiment					
from endline survey	0.12	0.07	-0.05	0.11	
	(0.02)	(0.01)	(0.03)		
from endline test	0.12	0.11	-0.01	0.83	
	(0.03)	(0.02)	(0.03)		
number of communities	27	26			

Panel A shows sample sizes by survey wave and treatment status. Original sample refers to those present at baseline. Added sample refers to new subjects added at endline who were not present at baseline. Interview refers to completed interview. Test score refers to completed MDAT test for fine motor skills and language/hearing. Panel B shows attrition rates by treatment group. Attrited from endline is indicator for not being present for endline interview, conditional on being present for baseline interview. Attrited from endline test is indicator for not being present for endline test, conditional on taking baseline test. Standard errors in parentheses, clustered by settlement. p-values adjusted for stratification of treatment by region.

Table 3: Baseline characteristics and balance tests

Experiment	Community-based ECD				ECD Annex			
Group	control	treatment	difference	p-value	Control	treatment	difference	p-value
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
age (months)	46.9	47.6	0.8	0.39	57.2	56.1	-1.1	0.75
	(0.4)	(0.5)	(0.7)		(1.4)	(1.3)	(1.9)	
female	0.49	0.44	-0.06	0.13	0.47	0.50	0.03	0.61
	(0.03)	(0.03)	(0.04)		(0.04)	(0.03)	(0.05)	
attend ECD	0.37	0.23	-0.14	0.04	0.88	0.92	0.05	0.47
	(0.05)	(0.04)	(0.06)		(0.04)	(0.03)	(0.05)	
fine motor skills	-0.12	-0.13	0.00	0.78	0.04	0.29	0.25	0.33
	(0.08)	(0.13)	(0.15)		(0.12)	(0.16)	(0.20)	
language and hearing skills	-0.22	-0.15	0.07	0.47	0.17	0.36	0.19	0.45
	(0.08)	(0.13)	(0.15)		(0.12)	(0.16)	(0.19)	
height-for-age	-1.35	-1.55	-0.20	0.11	-1.37	-1.28	0.09	0.64
	(0.08)	(0.09)	(0.12)		(0.12)	(0.21)	(0.24)	
household size	7.7	8.2	0.5	0.27	7.6	8.0	0.4	0.48
	(0.3)	(0.3)	(0.4)		(0.3)	(0.4)	(0.5)	
mother's schooling	1.8	1.7	0.0	0.67	1.7	2.2	0.5	0.60
	(0.3)	(0.3)	(0.4)		(0.2)	(0.4)	(0.4)	
household head schooling	2.2	1.8	-0.4	0.88	2.4	2.2	-0.2	0.13
	(0.3)	(0.3)	(0.5)		(0.4)	(0.4)	(0.6)	
household head employed	0.82	0.84	0.03	0.43	0.82	0.73	-0.09	0.10
	(0.03)	(0.03)	(0.05)		(0.04)	(0.05)	(0.07)	
household head work hours	38.0	37.8	-0.2	0.95	38.3	33.4	-4.9	0.23
	(2.5)	(2.4)	(3.4)		(2.7)	(3.3)	(4.3)	
household head in agriculture	0.57	0.72	0.15	0.05	0.60	0.55	-0.05	0.86
	(0.04)	(0.04)	(0.06)		(0.04)	(0.05)	(0.07)	
household expenditure per capita	483.2	523.6	40.4	0.56	519.0	418.4	-100.6	0.06
	(67.9)	(56.3)	(87.8)		(40.2)	(29.4)	(49.4)	
asset index	-0.04	-0.52	-0.48	0.05	-0.03	0.68	0.72	0.14
	(0.22)	(0.13)	(0.25)		(0.28)	(0.35)	(0.44)	

ECD willingness to pay	0.11	0.21	0.10	0.10	0.07	0.18	0.11	0.11
(as % of household expenditure)	(0.02)	(0.07)	(0.07)		(0.01)	(0.06)	(0.06)	
vaccinations (% of 17)	0.55	0.60	0.05	0.21	0.54	0.56	0.02	0.78
	(0.03)	(0.02)	(0.04)		(0.03)	(0.04)	(0.05)	
child ill	0.13	0.17	0.04	0.26	0.18	0.16	-0.02	0.53
	(0.02)	(0.02)	(0.03)		(0.03)	(0.03)	(0.04)	
mother mental distress	44.9	41.8	-3.1	0.22	47.6	45.8	-1.8	0.64
(0-100 scale, where 100 is worst)	(1.7)	(2.4)	(2.9)		(2.9)	(3.0)	(4.1)	
stimulating objects (% of 10)	0.27	0.28	0.00	0.85	0.28	0.30	0.01	0.58
	(0.01)	(0.01)	(0.02)		(0.01)	(0.01)	(0.02)	
corporal punishment (% use)	0.71	0.66	-0.05	0.38	0.63	0.69	0.07	0.51
	(0.03)	(0.04)	(0.05)		(0.05)	(0.05)	(0.06)	
Observations	319	267			192	182		

Sample is children with baseline MDAT score. All variables are means from baseline survey. Standard errors in parentheses, clustered by settlement. p-values adjusted for stratification of treatment by region. Fine motor, language and hearing skills are z-scores from MDAT (adjusted scores based on standardized residuals from regression of raw score on child's age, age squared, and female dummy). Height-for-age z-score based on World Health Organization 2007 benchmark. Household expenditure per capita is annual value in USD, winsorized at 1st/99th percentiles. Asset index is first principal component of reported household assets. ECD willingness to pay is stated willingness to pay for early childhood development services as share of household per capital expenditure. Vaccinations is proportion of 17 vaccinations received by child. Mother mental distress is percentage of 11 mental health issues experienced "most of the time" by mother. Stimulating objects is proportion of 10 objects for stimulating play found in home. Corporal punishment is indicator for using corporal punishment as usual form of discipline for severe misbehavior.

Table 4: Implementation measures

Variable	ECD Annex treated (1)	ECD Community-based (2)	difference (3)	p-value (4)
Site characteristics				
registered children (male)	24.3	22.3	0.9	0.78
	(2.39)	(2.08)	(3.25)	
registered children (female)	23.5	23.1	-0.5	0.90
	(2.18)	(2.72)	(3.81)	
registered children (total)	47.8	45.4	0.4	0.95
	(4.04)	(4.64)	(6.61)	
children present (male)	17.3	18.8	-2.2	0.44
	(1.80)	(1.94)	(2.80)	
children present (female)	15.6	19.7	-3.8	0.24
	(1.68)	(2.33)	(3.20)	
children present (total)	32.2	38.4	-6.6	0.25
	(3.25)	(4.00)	(5.65)	
attendance % (male)	0.72	0.85	-0.12	0.02
	(0.03)	(0.03)	(0.05)	
attendance % (female)	0.69	0.92	-0.18	0.05
	(0.04)	(0.07)	(0.09)	
attendance % (total)	0.71	0.88	-0.14	0.01
	(0.04)	(0.04)	(0.05)	
pupil-teacher ratio	35.7	31.4	-0.4	0.94
	(4.01)	(3.97)	(5.38)	
Teacher characteristics				
female	0.32	0.38	0.07	0.81
	(0.08)	(0.07)	(0.11)	
age	33.3	27.0	-6.3	0.00
	(1.3)	(0.9)	(1.6)	
at least primary	0.64	0.85	0.21	0.02
	(0.08)	(0.04)	(0.09)	
at least secondary	0.61	0.30	-0.31	0.00
	(0.07)	(0.06)	(0.09)	
qualified	0.50	0.03	-0.47	0.00
	(0.11)	(0.03)	(0.12)	
teaching experience	5.7	0.5	-5.1	0.00
	(1.3)	(0.2)	(1.3)	
Ever taught at ECD	0.54	0.13	-0.41	0.00
	(0.11)	(0.05)	(0.12)	
ECD experience	1.3	0.3	-1.0	0.01
	(0.3)	(0.1)	(0.4)	
Citizen of the village	0.16	0.65	0.49	0.00
	(0.08)	(0.07)	(0.10)	
commute time (minutes)	18.2	15.2	-3.0	0.43
	(3.4)	(2.4)	(4.2)	
work hours last week	18.5	17.8	-0.6	0.95
	(2.0)	(1.9)	(2.8)	
absent from class last month	0.42	0.21	-0.21	0.03
	(0.08)	(0.08)	(0.11)	
named theme from syllabus	0.57	0.43	-0.14	0.21
as this week's topic	(0.06)	(0.06)	(0.09)	
follows syllabus fully	0.43	0.37	-0.07	0.66
	(0.06)	(0.06)	(0.08)	
<u>To what extent are children…? Very much</u>				
…asking questions?	0.54	0.72	0.19	0.17
	(0.11)	(0.08)	(0.13)	
…taking initiative?	0.54	0.76	0.22	0.08
	(0.10)	(0.08)	(0.12)	

…interacting among themselves?	0.82	0.86	0.04	0.72
	(0.08)	(0.06)	(0.10)	
…listening and responding to teacher?	0.75	0.90	0.15	0.17
	(0.09)	(0.06)	(0.11)	
…interacting with toys and books?	0.89	0.90	0.01	0.94
	(0.06)	(0.06)	(0.08)	
Observations				
Teachers	44	60		
Sites	25	38		

Table shows results of monitoring survey of ECD providers in treatment sites, 2013. Teaching experience measured in years.

Table 5: Endline outcomes

Experiment	Community-based ECD				ECD Annex			
Group	control	treatment	treatment effect		control	treatment	treatment effect	
			without baseline outcome	with baseline outcome			without baseline outcome	with baseline outcome
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Language and hearing								
overall score (z)	-0.07	-0.25	-0.16	-0.17	0.12	0.30	0.14	0.11
	(0.08)	(0.10)	(0.12)	(0.12)	(0.10)	(0.12)	(0.15)	(0.15)
knows own name & its letters (% of 4)	0.99	0.98	-0.01	-0.01	1.00	1.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
speaks in clear sentences	0.99	0.97	-0.02	-0.04	0.99	0.99	0.00	0.00
	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
counting (% of 3)	0.39	0.43	0.04	0.04	0.49	0.45	-0.04	-0.09**
	(0.02)	(0.03)	(0.04)	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)
name colors (% of 4)	0.16	0.19	0.02	0.01	0.24	0.25	0.02	-0.01
	(0.01)	(0.02)	(0.03)	(0.03)	(0.01)	(0.02)	(0.02)	(0.03)
play with blocks (% of 3)	0.38	0.37	0.00	N/A	0.45	0.43	-0.02	N/A
	(0.02)	(0.02)	(0.03)		(0.02)	(0.02)	(0.03)	
open books (% of 3)	0.18	0.17	0.00	N/A	0.23	0.19	-0.04	N/A
	(0.01)	(0.02)	(0.02)		(0.02)	(0.02)	(0.03)	
Fine motor skills								
overall score (z)	-0.07	-0.15	-0.05	-0.04	0.06	0.24	0.11	0.16
	(0.10)	(0.12)	(0.12)	(0.14)	(0.10)	(0.12)	(0.14)	(0.14)
play with blocks (% of 6)	0.38	0.40	0.03	-0.01	0.42	0.40	-0.02	-0.02
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
draw lines & shapes (% of 6)	0.41	0.45	0.05	0.03	0.45	0.43	-0.02	-0.04
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
order rows of items (% of 2)	0.29	0.29	0.00	-0.01	0.39	0.34	-0.06	-0.08*
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.04)	(0.04)
fold paper	0.62	0.62	0.01	N/A	0.68	0.76	0.07	N/A
	(0.04)	(0.05)	(0.06)		(0.06)	(0.05)	(0.08)	
play with blocks	0.38	0.39	0.01	N/A	0.43	0.41	-0.02	N/A
(language and motors skills combined, % of 6)	(0.02)	(0.02)	(0.03)		(0.02)	(0.02)	(0.03)	
Observations								
children	345	281			207	188		
sites	50	40			27	26		

* significant at 10%, ** significant at 5%, *** significant at 1%. All variables are means from MDAT endline survey. Standard errors in parentheses, clustered by settlement. Treatment effects control for stratification of treatment assignment by region. Estimates in columns (4) and (8) augment regression with control for baseline outcome. z-scores are standardized residuals from regression of raw score on child's age, age squared, and female dummy. Other variables are subsets of items on MDAT test, measured as percent of items completed correctly. Speaks in clear sentences is just one item, while other categories have number of items indicated.

Table 6(a): Treatment effect heterogeneity, Community-based ECD experiment

Outcome	Fine motor skills						Language and hearing					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treatment interacted with:												
male	0.03						-0.10					
	(0.11)						(0.11)					
female	-0.09						-0.28*					
	(0.11)						(0.11)					
assets below median		-0.01						-0.13				
		(0.12)						(0.12)				
assets above median		-0.12						-0.40**				
		(0.13)						(0.13)				
mother didn't attend school			0.05						-0.12			
			(0.09)						(0.09)			
mother attended school			-0.32						-0.40*			
			(0.17)						(0.17)			
stimulating objects below median				0.14						-0.04		
				(0.10)						(0.10)		
stimulating objects above median				-0.24*						-0.36**		
				(0.12)						(0.12)		
child not sick last 3 days					-0.05						-0.20*	
					(0.08)						(0.08)	
child sick last 3 days					0.11						-0.01	
					(0.22)						(0.23)	
mother mental distress below median						-0.04						-0.24*
						(0.11)						(0.11)
mother mental distress above median						-0.09						-0.28*
						(0.14)						(0.14)
N	622						618					
H1: low SES treatment effect=0	0.59						0.59					
H2: high SES treatment effect=0	0.17						0.05					
H3: low SES = high SES	0.02						0.01					

* significant at 10%, ** significant at 5%, *** significant at 1%. Fine motor skills and language and hearing outcomes are MDAT z-scores, adjusted for age, age squared, and gender. Each column shows results of a separate regression. All regressions include control for Region 2 and dummies for all subgroups listed in table (constant omitted so that all subgroup coefficients and interactions identified). p-values reported at bottom from tests across equations, using joint variance-covariance matrix of all coefficients. "Low SES treatment effect=0" reports p-value of joint test on interactions between treatment and indicators for assets below median, mother didn't attend school, and stimulating objects below median. "High SES treatment effect=0" reports p-value of joint test on interactions between treatment and indicators for assets above median, mother attended school, and stimulating objects above median. "Low SES=High SES" reports p-value of joint test of equality between low-SES interactions and high-SES interactions. Household assets based on first principal component of asset ownership. Mother mental distress is proportion of responses to mental health questionnaire answering "3," which is "most of the time" for mental health issue. Stimulating objects in home based on proportion of 10 objects owned. Standard errors in parenthesis, clustered by settlement.

Table 6(b): Treatment effect heterogeneity, ECD Annex experiment

Outcome	Fine motor skills						Language and hearing					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treatment interacted with:												
male	0.24						0.26					
	(0.13)						(0.14)					
female	-0.03						-0.01					
	(0.14)						(0.15)					
assets below median		0.09						0.07				
		(0.15)						(0.16)				
assets above median		0.21						0.15				
		(0.13)						(0.15)				
mother didn't attend school			0.08						0.10			
			(0.11)						(0.12)			
mother attended school			0.23						0.22			
			(0.19)						(0.20)			
stimulating objects below median				0.05						-0.08		
				(0.13)						(0.14)		
stimulating objects above median				0.18						0.34*		
				(0.13)						(0.15)		
child not sick last 3 days					0.16						0.17	
					(0.10)						(0.11)	
child sick last 3 days					-0.10						-0.08	
					(0.24)						(0.26)	
mother mental distress below median						-0.04						0.00
						(0.13)						(0.15)
mother mental distress above median						0.39**						0.26
						(0.15)						(0.16)
N	392						388					
H1: low SES treatment effect=0	0.95						0.39					
H2: high SES treatment effect=0	0.66						0.35					
H3: low SES = high SES	0.31						0.16					

* significant at 10%, ** significant at 5%, *** significant at 1%. Fine motor skills and language and hearing outcomes are MDAT z-scores, adjusted for age, age squared, and gender. Each column shows results of a separate regression. All regressions include control for Region 2 and dummies for all subgroups listed in table (constant omitted so that all subgroup coefficients and interactions identified). p-values reported at bottom from tests across equations, using joint variance-covariance matrix of all coefficients. "Low SES treatment effect=0" reports p-value of joint test on interactions between treatment and indicators for assets below median, mother didn't attend school, and stimulating objects below median. "High SES treatment effect=0" reports p-value of joint test on interactions between treatment and indicators for assets above median, mother attended school, and stimulating objects above median. "Low SES=High SES" reports p-value of joint test of equality between low-SES interactions and high-SES interactions. Household assets based on first principal component of asset ownership. Mother mental distress is proportion of responses to mental health questionnaire answering "3," which is "most of the time" for mental health issue. Stimulating objects in home based on proportion of 10 objects owned. Standard errors in parenthesis, clustered by settlement.

Table 7: Home investments (Time)

	Outcome: mother play time with child (minutes/day)													
Experiment	Community-based ECD							ECD Annex						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Treatment	5.6							46.2						
	(19.4)							(27.9)						
Treatment interacted with:														
male		-8.1							54.6**					
		(15.6)							(17.8)					
female		20.0							36.9*					
		(16.9)							(18.6)					
assets below median			12.0							25.8				
			(17.5)							(20.8)				
assets above median			13.5							56.5**				
			(19.0)							(18.3)				
mother didn't attend school				1.4							56.7***			
				(12.9)							(14.6)			
mother attended school				18.2							9.3			
				(25.4)							(27.0)			
stimulating objects below median					-2.0							30.3		
					(15.2)							(18.8)		
stimulating objects above median					14.0							61.0***		
					(17.5)							(18.2)		
child not sick last 3 days						10.7							36.2**	
						(12.3)							(14.0)	
child sick last 3 days						-34.8							104.1**	
						(32.0)							(33.4)	
mother mental distress below median							3.3							43.2*
							(16.5)							(18.2)
mother mental distress above median							25.6							41.0
							(20.5)							(21.1)
N	1,045	1,045	1,045	1,045	1,045	1,045	1,045	728	728	728	728	728	728	728
Control group mean	184.5							162.4						
H1: low SES treatment effect=0				0.89							0.02			
H2: high SES treatment effect=0				0.92							0.04			
H3: low SES = high SES				0.50							0.78			

* significant at 10%, ** significant at 5%, *** significant at 1%. Outcome is mother's play time with child (minutes/day). Each column shows results of a separate regression. All regressions include control for Region 2. Regressions with interacted treatment variables include dummies for main effect of all subgroups listed in table (constant omitted so that all subgroup coefficients and interactions identified). p-values reported at bottom from tests across equations, using joint variance-covariance matrix of all coefficients. "Low SES treatment effect=0" reports p-value of joint test on interactions between treatment and indicators for assets below median, mother didn't attend school, and stimulating objects below median. "High SES treatment effect=0" reports p-value of joint test on interactions between treatment and indicators for assets above median, mother attended school, and stimulating objects above median. "Low SES=High SES" reports p-value of joint test of equality between low-SES interactions and high-SES interactions. Household assets based on first principal component of asset ownership. ECD demand is stated willingness to pay for ECD services as percentage of household per capita expenditure. Mother mental distress is proportion of responses to mental health questionnaire answering "3," which is "most of the time" for mental health issue. Stimulating objects in home based on proportion of 10 objects owned. Standard errors in parenthesis, clustered by settlement.

Table 8: Home investments (Stimulation Items)

	Outcome: stimulating items in household (proportion of 11)													
Experiment	Community-based ECD							ECD Annex						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Treatment	0.03							-0.02						
	(0.04)							(0.04)						
Treatment interacted with:														
male		0.05**							-0.02					
		(0.02)							(0.03)					
female		0.00							-0.04					
		(0.02)							(0.03)					
assets below median			0.05*							-0.04				
			(0.02)							(0.03)				
assets above median			-0.01							-0.04				
			(0.02)							(0.03)				
mother didn't attend school				0.03							-0.02			
				(0.02)							(0.02)			
mother attended school				0.01							-0.05			
				(0.03)							(0.04)			
stimulating objects below median					0.03							-0.02		
					(0.02)							(0.03)		
stimulating objects above median					0.03							-0.04		
					(0.02)							(0.03)		
child not sick last 3 days						0.02							-0.03	
						(0.02)							(0.02)	
child sick last 3 days						0.05							0.00	
						(0.04)							(0.05)	
mother mental distress below median							0.03							-0.04
							(0.02)							(0.03)
mother mental distress above median							0.00							-0.04
							(0.03)							(0.03)
N	1,055	1,055	1,055	1,055	1,055	1,055	1,055	730	730	730	730	730	730	730
Control group mean	0.30							0.36						
H1: low SES treatment effect=0				0.61							0.75			
H2: high SES treatment effect=0				0.57							0.74			
H3: low SES = high SES				0.35							0.66			

* significant at 10%, ** significant at 5%, *** significant at 1%. Outcome is simulating objects found in home (proportion of list of 11). Each column shows results of a separate regression. All regressions include control for Region 2. Regressions with interacted treatment variables include dummies for main effect of all subgroups listed in table (constant omitted so that all subgroup coefficients and interactions identified). p-values reported at bottom from tests across equations, using joint variance-covariance matrix of all coefficients. "Low SES treatment effect=0" reports p-value of joint test on interactions between treatment and indicators for assets below median, mother didn't attend school, and stimulating objects below median. "High SES treatment effect=0" reports p-value of joint test on interactions between treatment and indicators for assets above median, mother attended school, and stimulating objects above median. "Low SES=High SES" reports p-value of joint test of equality between low-SES interactions and high-SES interactions. Household assets based on first principal component of asset ownership. ECD demand is stated willingness to pay for ECD services as percentage of household per capita expenditure. Mother mental distress is proportion of responses to mental health questionnaire answering "3," which is "most of the time" for mental health issue. Stimulating objects in home based on proportion of 10 objects owned. Standard errors in parenthesis, clustered by settlement.

Appendix: Estimating the relative effectiveness of ECD Community Centers and ECD Annexes

Method

As discussed in the introduction, meeting the twin goals of access and quality in early childhood development services forces governments to make difficult choices. Comparing the efficacy of the community-based ECD treatment, which attempts to expand access to ECD services, with that of the ECD Annex treatment, which aims to increase quality, matters greatly for policy. In addition, it is important for our paper to establish whether there are important quality differences between these two types of center-based care. However, differences in the populations eligible for the two experiments present challenges for this comparison, as noted earlier.

We attempt to overcome this challenge by comparing *ex ante* similar children across experiments. Specifically, we compare the experimental treatments against each other using an inverse propensity score weighting estimator, which combines matching and reweighting to improve comparability between treated and control units (Hirano, Imbens, and Ridder 2003; Abadie 2005).[21] We first match on the propensity score by estimating the probability that the child is in the ECD Annex experiment and not the community-based ECD experiment (regardless of treatment status within each experiment).[22] We then remove children falling outside the common support of the propensity score distributions for each experiment.

Using this trimmed sample, we weight by the inverse propensity score and estimate:

$$y_{ic} = \alpha + \beta_1 community_c + \beta_2 Annex_c + \gamma Region2_c + \varepsilon_{ic} \quad \text{(B.1)}$$

where *community* is an indicator for being in the community-based ECD treatment group, *Annex* is an indicator for being in the ECD Annex experiment, and all else is as in equation (1). Because the community-based ECD control group is the omitted category, the coefficients β_1 and β_2 measure the treatment effect of being in each of the corresponding groups relative to the pure control.

[21] The exposition in this section closely follows Giordono and Pugatch (2017).
[22] Section 5 presents details of propensity score estimation.

The identifying assumption is that treatment assignment is orthogonal to unobserved characteristics that also affect the outcome. In the case of the community-based ECD treatment, this is ensured by random assignment, whereas in the case of the ECD Annex experiment we rely on the weights to balance average characteristics with the community-based ECD sample. Although the method requires that unobserved characteristics do not systematically differ among children in different experimental groups with the same propensity score, we rely on a rich set of predetermined characteristics, including baseline MDAT scores, to generate the weights. The next section presents details of propensity score estimation. Moreover, the method consistently estimates the average treatment effect even if the propensity score equation is misspecified (Robins and Rotnitzky 1995; Imbens and Wooldridge 2009). Under this identifying assumption, we can also compare the ECD Annex experiment to the community-based ECD treatment through a test of the null hypothesis $\beta_2 - \beta_1 = 0$.

Finally, we can compare the ECD Annex control and treatment separately to the community-based ECD groups by disaggregating the *Annex* experiment dummy in equation (B.1) into separate indicators of treatment assignment:

$$y_{ic} = \alpha + \beta_1 community_c + \beta_2 Annex\_control_c + \beta_3 Annex\_treated_c + \gamma Region2_c + \varepsilon_{ic} \quad \text{(B.2)}$$

Estimates of β_2 and β_3 in equation (B.2) help to determine if the Annex setting without training leads to differences with the community-based experiment, or whether the provider training is essential for implementing the new curriculum.

Results

The populations eligible for each experiment live in distinct communities. Villages in the community-based ECD experiment lacked access to structured ECD services prior to the treatment, whereas villages in the ECD Annex experiment already had an ECD facility operating in conjunction with a primary school. These conditions reflect broader differences between households in the two experiments.

To ensure comparability between the two experimental samples, we first estimate the probability that a child is in the ECD Annex sample by running a logit regression on a host of baseline characteristics. These characteristics are the child's age (in months); dummies for

female, region, and previous ECD attendance; baseline z-scores in fine motor skills, language and hearing, and height-for-age; mother's schooling (in years); household expenditure per capita and willingness to pay for ECD services as a percentage of this expenditure; the proportion of 17 recommended vaccines received by the child; and an index of mother's mental distress, measured as the proportion of 11 mental health issues she reports experiencing "most of the time." We then exclude children whose propensity score falls outside the common support of the distributions from each experiment, leaving an estimation sample that includes 648 of the 844 children (77%) with an endline MDAT score.[23]

In Table A4, we check for balance in baseline fine motor skills and language and hearing using the trimmed sample. MDAT scores differ significantly between experiments when the data are unweighted, for both overall scores on each module and for most subsets of test items (columns 1-4). When weighting observations by the inverse of their propensity score (columns 5-8), no score differs at conventional significance levels.[24] These results give us confidence that reweighting generates an appropriate sample with which to compare experimental treatments.

In Table A5, we present results from inverse propensity score-weighted estimation of the effect of each experimental treatment. Column 1 shows estimates of equation (B.1) for language and hearing skills. The coefficient on the community-based ECD treatment indicator corresponds to a decrease of 0.64 standard deviations in language skills relative to the pure control group (the omitted category), significant at 5%. The effect attenuates somewhat to -0.37 when controlling for baseline score in column 2, but remains significant. These results are striking, given that the pure control group lacks ready access to structured ECD services and the variation between these groups is experimental. The estimates also contrast with the previous estimates of the community-based ECD treatment effect of -0.16 and -0.17 standard deviations, which were not statistically distinguishable from zero (Table 5, columns 3-4).

A clue to this discrepancy lies in the pure control group mean language score of 0.49 (reported halfway down Table A5), which is considerably higher than the unweighted mean of -0.07 (Table 5, column 1). In other words, the reweighting used to estimate equation (B.1) gives

[23] Figure SA2 of the Supplemental Appendix shows the propensity score distributions.

[24] The result is not simply a mechanical consequence of reweighting. Although the overall score on each module enters the model for the propensity score, it is not obvious that the procedure would also succeed in balancing the means of each subset of items.

greater influence to relatively better-off children, who fare better at home than in community-based ECD. This interpretation also echoes the subgroup analysis of Table 6, in which we found the same pattern among more advantaged children in the community-based ECD experiment. The results also match broader lessons from the literature on the heterogeneous effects of pre-school by socioeconomic status.

Returning to column 1 of Table A5, the coefficient on the pooled ECD Annex group is -0.17, but not statistically significant, indicating no distinguishable difference in outcomes between observationally equivalent children in pure control and ECD Annex communities. However, a test of the difference between the ECD Annex and community-based ECD coefficients, reported at the bottom of the table, shows an advantage of 0.47 standard deviations for the ECD Annex group, significant at 1%. The Lee (2009) bounds for this estimate (reported in brackets) exclude zero, alleviating concerns that differential attrition between experiments drives the result.[25] This effect falls to 0.40 standard deviations and remains significant at 10% when controlling for baseline score in column 2. These results suggest that providing ECD services through annexes attached to primary schools is more effective at promoting language development than community-based ECD services, at least during our sample period.

Table A5, columns 3 and 4 disaggregate the ECD Annex group by treatment and control, corresponding to equation (B.2). The ECD Annex treatment and control have nearly identical effects relative to both the pure control group (as shown by their coefficients) and to the community-based ECD treatment (as shown by the estimates reported at the bottom of the table). Bounds for these estimates are also nearly identical. These results suggest that the teacher training provided in the ECD Annex treatment added little value relative to the control, at least by the end of the sample period. Instead, it appears that other features of the ECD Annex environment explain its differences with the community-based ECD program.

Columns 5 to 8 of Table A5 repeat the exercise for fine motor skills. For this outcome, we find no statistically significant differences between the pure control group and the community-based ECD treatment. Nor are there significant differences between the pure control

[25] The Lee (2009) bounds apply to a pairwise comparison between groups, but continue to weight by the inverse propensity score and adjust for regional stratification as in the regressions. We cannot bound the effect while controlling for the baseline outcome because there must be variation in treatment within covariate cells, whereas normalized MDAT scores are continuous.

and the ECD Annex sample, either pooled or separately by treatment status. Children in the ECD Annex group do score significant higher than those in the community-based treatment, however, with magnitudes ranging from 0.35-0.42 standard deviations, though the effects lose significance when controlling for baseline outcome. As with language and hearing, for fine motor skills we again see no discernible benefit to the ECD Annex treatment relative to the control.

Table A1: Endline outcomes, with controls for baseline imbalance

Experiment	Community-based ECD				ECD Annex			
Group	control	treatment	treatment effect		control	treatment	treatment effect	
			controls for imbalance	controls + baseline outcome			controls for imbalance	controls + baseline outcome
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Language and hearing								
overall score (z)	-0.07	-0.25	-0.09	-0.09	0.12	0.30	0.12	0.11
	(0.08)	(0.10)	(0.11)	(0.11)	(0.10)	(0.12)	(0.15)	(0.16)
knows own name & its letters (% of 4)	0.99	0.98	-0.01	-0.01	1.00	1.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
speaks in clear sentences	0.99	0.97	-0.02	-0.03	0.99	1.00	0.01	0.01
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
counting (% of 3)	0.70	0.68	0.01	0.00	0.87	0.88	0.00	-0.02
	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)
name colors (% of 4)	0.29	0.29	0.02	0.01	0.42	0.49	0.08*	0.03
	(0.02)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)
play with blocks (% of 3)	0.67	0.59	-0.07*	N/A	0.79	0.83	0.03	N/A
	(0.03)	(0.02)	(0.03)		(0.03)	(0.03)	(0.04)	
open books (% of 3)	0.32	0.27	-0.02	N/A	0.40	0.37	-0.03	N/A
	(0.02)	(0.03)	(0.03)		(0.04)	(0.04)	(0.06)	
Fine motor skills								
overall score (z)	-0.07	-0.15	0.02	0.04	0.06	0.24	0.08	0.13
	(0.10)	(0.12)	(0.12)	(0.13)	(0.10)	(0.12)	(0.13)	(0.14)
play with blocks (% of 6)	0.68	0.63	-0.02	-0.02	0.73	0.78	0.03	0.03
	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
draw lines & shapes (% of 6)	0.72	0.71	0.01	0.01	0.79	0.83	0.02	0.02
	(0.02)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
order rows of items (% of 2)	0.52	0.45	-0.05	-0.05	0.69	0.65	-0.05	-0.05
	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.05)	(0.05)
fold paper	0.62	0.62	0.03	N/A	0.68	0.76	0.07	N/A
	(0.04)	(0.05)	(0.05)		(0.06)	(0.05)	(0.07)	
play with blocks	0.68	0.61	-0.04*	N/A	0.76	0.80	0.03	N/A
(language and motors skills combined, % of 6)	(0.02)	(0.02)	(0.03)		(0.03)	(0.02)	(0.03)	
Observations								
children	345	281			207	188		
sites	50	40			27	26		

* significant at 10%, ** significant at 5%, *** significant at 1%. All variables are means from MDAT endline survey. Standard errors in parentheses, clustered by settlement. All regressions include dummy for Region 2 and variables imbalanced at baseline (household asset index, household expenditure per capita, and dummies baseline ECD attendance and household head in agriculture). Baseline outcome included where indicated. Overall scores are standardized residuals from regression of raw score on child's age, age squared, and female dummy. Other variables are subsets of items on MDAT test, measured as percent of items completed correctly. Speaks in clear sentences is just one item, while other categories have number of items indicated.

Table A2(a): Treatment effect heterogeneity, including baseline ECD attendance and imbalanced variables, Community-based ECD experiment

Outcome	Fine motor skills							Language and hearing						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Treatment interacted with:														
male	0.14							-0.07						
	(0.12)							(0.12)						
female	-0.10							-0.24						
	(0.13)							(0.13)						
assets below median		0.08							-0.05					
		(0.12)							(0.12)					
assets above median		-0.05							-0.27*					
		(0.13)							(0.13)					
mother didn't attend school			0.12							-0.07				
			(0.10)							(0.10)				
mother attended school			-0.25							-0.37*				
			(0.18)							(0.18)				
stimulating objects below median				0.20							-0.03			
				(0.13)							(0.13)			
stimulating objects above median				-0.12							-0.25*			
				(0.12)							(0.12)			
child not sick last 3 days					0.00							-0.19*		
					(0.10)							(0.10)		
child sick last 3 days					0.19							0.12		
					(0.22)							(0.23)		
mother mental distress below median						0.06							-0.10	
						(0.11)							(0.11)	
mother mental distress above median						-0.02							-0.21	
						(0.14)							(0.14)	
attended ECD at baseline							0.01							-0.11
							(0.16)							(0.16)
did not attend ECD at baseline							0.03							-0.16
							(0.11)							(0.11)
N				476							472			
H1: low SES treatment effect=0				0.62							0.92			
H2: high SES treatment effect=0				0.36							0.14			
H3: low SES = high SES				0.03							0.03			

* significant at 10%, ** significant at 5%, *** significant at 1%. Fine motor skills and language and hearing outcomes are MDAT z-scores, adjusted for age, age squared, and gender. Each column shows results of a separate regression. All regressions include control for Region 2, dummies for all subgroups listed in table (constant omitted so that all subgroup coefficients and interactions identified), and variables imbalanced at baseline (ECD attendance, household expenditure per capita, household assets, and household head in agriculture). p-values reported at bottom from tests across equations, using joint variance-covariance matrix of all coefficients. "Low SES treatment effect=0" reports p-value of joint test on interactions between treatment and indicators for assets below median, mother didn't attend school, and stimulating objects below median. "High SES treatment effect=0" reports p-value of joint test on interactions between treatment and indicators for assets above median, mother attended school, and stimulating objects above median. "Low SES=High SES" reports p-value of joint test of equality between low-SES interactions and high-SES interactions. Household assets based on first principal component of asset ownership. ECD demand is stated willingness to pay for ECD services as percentage of household per capita expenditure. Mother mental distress is proportion of responses to mental health questionnaire answering "3," which is "most of the time" for mental health issue. Stimulating objects in home based on proportion of 10 objects owned. Standard errors in parenthesis, clustered by settlement.

Table A2(b): Treatment effect heterogeneity, including baseline ECD attendance and imbalanced variables, ECD Annex experiment

Outcome	Fine motor skills							Language and hearing						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Treatment interacted with:														
male	0.17							0.15						
	(0.15)							(0.16)						
female	0.03							0.04						
	(0.16)							(0.18)						
assets below median		0.13							0.15					
		(0.17)							(0.18)					
assets above median		0.09							0.06					
		(0.15)							(0.17)					
mother didn't attend school			0.04							0.04				
			(0.13)							(0.14)				
mother attended school			0.27							0.24				
			(0.20)							(0.23)				
stimulating objects below median				0.00							-0.24			
				(0.17)							(0.19)			
stimulating objects above median				0.19							0.36*			
				(0.15)							(0.16)			
child not sick last 3 days					0.17							0.14		
					(0.12)							(0.14)		
child sick last 3 days					-0.14							-0.07		
					(0.25)							(0.28)		
mother mental distress below median						-0.12							0.03	
						(0.15)							(0.16)	
mother mental distress above median						0.39*							0.19	
						(0.17)							(0.19)	
attended ECD at baseline							0.09							0.12
							(0.12)							(0.13)
did not attend ECD at baseline							0.25							-0.10
							(0.36)							(0.39)
N				295							291			
H1: low SES treatment effect=0				0.91							0.21			
H2: high SES treatment effect=0				0.53							0.21			
H3: low SES = high SES				0.39							0.11			

* significant at 10%, ** significant at 5%, *** significant at 1%. Fine motor skills and language and hearing outcomes are MDAT z-scores, adjusted for age, age squared, and gender. Each column shows results of a separate regression. All regressions include control for Region 2, dummies for all subgroups listed in table (constant omitted so that all subgroup coefficients and interactions identified), and variables imbalanced at baseline (ECD attendance, household expenditure per capita, household assets, and household head in agriculture). p-values reported at bottom from tests across equations, using joint variance-covariance matrix of all coefficients. "Low SES treatment effect=0" reports p-value of joint test on interactions between treatment and indicators for assets below median, mother didn't attend school, and stimulating objects below median. "High SES treatment effect=0" reports p-value of joint test on interactions between treatment and indicators for assets above median, mother attended school, and stimulating objects above median. "Low SES=High SES" reports p-value of joint test of equality between low-SES interactions and high-SES interactions. Household assets based on first principal component of asset ownership. ECD demand is stated willingness to pay for ECD services as percentage of household per capita expenditure. Mother mental distress is proportion of responses to mental health questionnaire answering "3," which is "most of the time" for mental health issue. Stimulating objects in home based on proportion of 10 objects owned. Standard errors in parenthesis, clustered by settlement.

Table A3(a): Treatment effect heterogeneity, Community-based ECD experiment, subsample not attending ECD at baseline

Outcome	Fine motor skills						Language and hearing					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treatment interacted with:												
male	0.14						-0.04					
	(0.15)						(0.14)					
female	-0.19						-0.43**					
	(0.16)						(0.16)					
assets below median		0.15						-0.10				
		(0.14)						(0.14)				
assets above median		-0.23						-0.37*				
		(0.16)						(0.16)				
mother didn't attend school			0.08						-0.16			
			(0.12)						(0.12)			
mother attended school			-0.36						-0.42			
			(0.24)						(0.24)			
stimulating objects below median				0.24						-0.09		
				(0.15)						(0.15)		
stimulating objects above median				-0.26						-0.34*		
				(0.15)						(0.15)		
child not sick last 3 days					-0.04						-0.29*	
					(0.12)						(0.11)	
child sick last 3 days					0.18						0.28	
					(0.28)						(0.28)	
mother mental distress below median						0.10						-0.14
						(0.14)						(0.14)
mother mental distress above median						-0.18						-0.32
						(0.17)						(0.17)
N	322						319					
H1: low SES treatment effect=0	0.60						0.73					
H2: high SES treatment effect=0	0.28						0.25					
H3: low SES = high SES	0.00						0.11					

* significant at 10%, ** significant at 5%, *** significant at 1%. Sample is children who did not attend ECD at baseline. Fine motor skills and language and hearing outcomes are MDAT z-scores, adjusted for age, age squared, and gender. Each column shows results of a separate regression. All regressions include control for Region 2 and dummies for all subgroups listed in table (constant omitted so that all subgroup coefficients and interactions identified). p-values reported at bottom from tests across equations, using joint variance-covariance matrix of all coefficients. "Low SES treatment effect=0" reports p-value of joint test on interactions between treatment and indicators for assets below median, mother didn't attend school, and stimulating objects below median. "High SES treatment effect=0" reports p-value of joint test on interactions between treatment and indicators for assets above median, mother attended school, and stimulating objects above median. "Low SES=High SES" reports p-value of joint test of equality between low-SES interactions and high-SES interactions. Household assets based on first principal component of asset ownership. Mother mental distress is proportion of responses to mental health questionnaire answering "3," which is "most of the time" for mental health issue. Stimulating objects in home based on proportion of 10 objects owned. Standard errors in parenthesis, clustered by settlement.

Table A3(b): Treatment effect heterogeneity, Community-based ECD experiment, subsample attending ECD at baseline

Outcome	Fine motor skills						Language and hearing					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treatment interacted with:												
male	-0.09						-0.33					
	(0.21)						(0.22)					
female	0.06						0.01					
	(0.24)						(0.25)					
assets below median		-0.16						-0.02				
		(0.23)						(0.24)				
assets above median		0.08						-0.33				
		(0.21)						(0.23)				
mother didn't attend school			0.09						0.00			
			(0.19)						(0.20)			
mother attended school			-0.26						-0.53			
			(0.27)						(0.28)			
stimulating objects below median				0.01						0.03		
				(0.26)						(0.26)		
stimulating objects above median				-0.05						-0.32		
				(0.20)						(0.21)		
child not sick last 3 days					-0.02						-0.17	
					(0.17)						(0.18)	
child sick last 3 days					-0.05						-0.25	
					(0.39)						(0.41)	
mother mental distress below median						-0.10						-0.18
						(0.21)						(0.21)
mother mental distress above median						0.07						-0.19
						(0.24)						(0.25)
N				154						153		
H1: low SES treatment effect=0				0.60						1.00		
H2: high SES treatment effect=0				0.54						0.12		
H3: low SES = high SES				0.75						0.03		

* significant at 10%, ** significant at 5%, *** significant at 1%. Sample is children who attended ECD at baseline. Fine motor skills and language and hearing outcomes are MDAT z-scores, adjusted for age, age squared, and gender. Each column shows results of a separate regression. All regressions include control for Region 2 and dummies for all subgroups listed in table (constant omitted so that all subgroup coefficients and interactions identified). p-values reported at bottom from tests across equations, using joint variance-covariance matrix of all coefficients. "Low SES treatment effect=0" reports p-value of joint test on interactions between treatment and indicators for assets below median, mother didn't attend school, and stimulating objects below median. "High SES treatment effect=0" reports p-value of joint test on interactions between treatment and indicators for assets above median, mother attended school, and stimulating objects above median. "Low SES=High SES" reports p-value of joint test of equality between low-SES interactions and high-SES interactions. Household assets based on first principal component of asset ownership. Mother mental distress is proportion of responses to mental health questionnaire answering "3," which is "most of the time" for mental health issue. Stimulating objects in home based on proportion of 10 objects owned. Standard errors in parenthesis, clustered by settlement.

Table A4: Between-experiments baseline balance tests

	unweighted				weighted			
	community-based	ECD Annex	(1) vs. (2)	p-value	community-based	ECD Annex	(1) vs. (2)	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Language and hearing								
overall score (z)	-0.19	0.27	0.58	0.00	0.33	0.09	-0.24	0.30
	(0.07)	(0.10)	(0.12)		(0.20)	(0.11)	(0.23)	
knows own name	0.13	0.19	0.06	0.00	0.18	0.17	-0.01	0.87
& its letters (% of 4)	(0.01)	(0.01)	(0.02)		(0.04)	(0.01)	(0.04)	
speaks in clear sentences	0.93	0.96	0.02	0.26	0.92	0.95	0.03	0.36
	(0.02)	(0.01)	(0.02)		(0.03)	(0.02)	(0.03)	
counting (% of 3)	0.09	0.25	0.17	0.00	0.23	0.19	-0.03	0.55
	(0.01)	(0.02)	(0.02)		(0.05)	(0.02)	(0.06)	
name colors (% of 4)	0.04	0.10	0.06	0.00	0.13	0.07	-0.07	0.15
	(0.01)	(0.02)	(0.02)		(0.04)	(0.01)	(0.05)	
Fine motor skills								
overall score (z)	-0.15	0.16	0.46	0.00	0.26	0.03	-0.23	0.33
	(0.08)	(0.10)	(0.13)		(0.22)	(0.11)	(0.24)	
play with blocks (% of 6)	0.35	0.36	0.03	0.11	0.41	0.36	-0.05	0.16
	(0.02)	(0.02)	(0.02)		(0.03)	(0.02)	(0.04)	
draw lines & shapes (% of 6)	0.35	0.40	0.07	0.01	0.44	0.39	-0.05	0.28
	(0.02)	(0.02)	(0.03)		(0.04)	(0.02)	(0.05)	
order rows of items (% of 2)	0.09	0.12	0.06	0.00	0.13	0.11	-0.02	0.67
	(0.01)	(0.02)	(0.02)		(0.06)	(0.02)	(0.06)	
Observations								
Children	328	320			328	320		
Sites	55	50			55	50		

All variables are means from baseline survey. Drops observations outside common support of propensity score distribution. Columns (5)-(8) weighted by inverse propensity score where indicated. Standard errors in parentheses, clustered by settlement. p-values obtained from regression of characteristic on community-based treatment and Region 2 dummy in order to adjust for stratification by region. Fine motor, language and hearing skills are z-scores from MDAT. Adjusted scores are standardized residuals from regression of raw score on child's age, age squared, and female dummy. Other variables are subsets of items on MDAT test, measured as percent of items completed correctly. Speaks in clear sentences is just one item, while other categories have number of items indicated

Table A5: Endline outcomes, combined experimental groups

MDAT module	Language and hearing				Fine motor			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Community-based ECD	-0.64	-0.37	-0.64	-0.36	-0.33	-0.14	-0.33	-0.14
	(0.28)**	(0.18)**	(0.28)**	(0.18)**	(0.25)	(0.21)	(0.25)	(0.21)
ECD Annex	-0.17	0.04			0.06	0.20		
	(0.29)	(0.22)			(0.21)	(0.16)		
ECD Annex control			-0.15	0.13			0.09	0.26
			(0.35)	(0.31)			(0.25)	(0.22)
ECD Annex treatment			-0.19	-0.07			0.02	0.12
			(0.27)	(0.20)			(0.21)	(0.15)
Observations								
Children	639	572	639	572	644	590	644	590
Sites	132	127	132	127	132	128	132	128
pure control group mean	0.49	0.49	0.49	0.49	0.23	0.23	0.23	0.23
	(1.14)	(1.14)	(1.14)	(1.14)	(0.99)	(0.99)	(0.99)	(0.99)
Includes baseline outcome		x		x		x		x
Difference from Community-based ECD								
ECD Annex	0.47***	0.40*			0.39**	0.34		
	(0.18)	(0.21)			(0.19)	(0.21)		
	[0.22,0.74]				[0.16,0.66]			
ECD Annex control			0.49*	0.49			0.42*	0.40
			(0.26)	(0.31)			(0.23)	(0.25)
			[0.15,0.74]				[0.42,0.76]	
ECD Annex treatment			0.45***	0.29*			0.35*	0.26
			(0.17)	(0.17)			(0.19)	(0.19)
			[0.15,0.68]				[0.11,0.58]	

* significant at 10%, ** significant at 5%, *** significant at 1%. Pure control group (i.e., control group from community-based ECD experiment) is omitted category. All regressions include region 2 dummy to adjust for stratification of treatment assignment and weight by inverse propensity score. Sample drops observations outside common support of propensity score distribution. Standard errors in parentheses, clustered by settlement. Outcomes are adjusted z-scores from MDAT modules for language and hearing and fine motor skills. Adjusted scores are standardized residuals from regression of raw score on child's age, age squared, and female dummy. Regressions include baseline outcome where indicated. Differences with community-based ECD reported at bottom of table based on tests of indicated coefficient with community-based ECD. Lee bounds reported in brackets, based on pairwise comparison, but still reweighting by inverse propensity score and stratifying by region. Propensity score obtained from logit model of membership in ECD Annex sample regressed on baseline characteristics. Included baseline characteristics: age (exact based on DOB), female, Region 2, ECD attendance, fine motor skills (age-adjusted z-score), language and hearing (age-adjusted z-score), height-for-age, household size, mother's years of schooling, household expenditure per capita (winsorized at 1st/99th percentiles), willingness to pay for ECD as % of household expenditure per capita, % of vaccines received, mother mental distress (% of items reported as experiencing "most of the time"). Missing values imputed to zero, with dummies for imputed value included as additional covariates in regression.