Closing the Early Childhood Skills Gap for Minorities: Evidence from Parental Investments and Preschool Attendance in Concentrated Roma Communities in Bulgaria*

Paul Gertler
Elise Huillery
Joost de Laat
Ricardo Miranda

February 2024 - PRELIMINARY, NOT FOR CITATION

Abstract

We estimate the relative contributions of preschool and parents to child development in disadvantaged communities in Bulgaria. Using a randomized control trial, we show that increased preschool attendance led to lower investments in parent engagement activities and lower child development. We estimate a child development production function, which reveals that parental investments and preschool attendance both contribute positively to child development but the contribution of parents is at least as large as the contribution of preschool, even for low-educated minority parents. These findings should encourage governments to recognize the strong contribution of parents and favor combinations that encourage preschool participation and parental stimulation at home, also when parents are extremely disadvantaged.

Keywords: Early child development, Preschool education, Parental investment, Human capital production function, Randomized control trial, Structural model

JEL classification codes: C93, I21, I24, I26, I28, J13, J24, 015

*Gertler: University of California at Berkeley, gertler@berkeley.edu; Huillery: Université Paris Dauphine – PSL, elise.huillery@dauphine.psl.eu; de Laat: University of Utrecht, j.j.delaat@uu.nl; Miranda Montero: University of Duke, r.e.mirandamontero@gmail.com. We thank the Trust for Social Achievement for their collaboration throughout the project, in particular Sarah Perrine and Eugenia Volen for outstanding project management, the staff of the twenty-three NGOs who implemented the interventions, and the staff of the Open Society Institute for excellent data collection. We are grateful to Robin Audy for outstanding research assistance. The funding for this study was provided by the World Bank's Srategic Impact Evaluation Fund (SIEF) (Call for projects FP7-SSH-2013-2). Plamen Djankov and Dessislava Kuznetsova led the implementation of this project on behalf of the World Bank, and we are deeply indebted to them for their leadership. All errors remain our own, and the opinions expressed in this paper are ours alone and should not be attributed to the institutions with which we are affiliated, the World Bank, TSA, or the Government of Bulgaria. This study is registered in the AEA RCT Registry and the unique identifying number is AEARCTR-XXX.

1 Introduction

Large socioeconomic disparities in cognitive development are evident when children enter primary school. Data from across the globe show important gaps between socio-economically advantaged and disadvantaged children at early stages, in both developed (Pillas et al. 2014; Bradley et al. 2002; Kelly et al. 2011; Waldfogel et al. 2011) and developing countries (Lu et al. 2020; López Boo 2016; Schady et al. 2015). The literature provides evidence that these disparities emerge within the two first years of life (Fernald et al. 2012) and widen between ages 3 and 5 (Kelly et al. 2011). Parental home stimulation and resources play an important role (Fernald et al. 2012; Akee et al. 2010) but these gradients remain even when accounting for nutrition and parenting (Naudeau et al. 2011). Research on inequality in early childhood development also shows that the development penalty for the disadvantaged children does not dissipate once they enter school (Schady et al. 2015) but leads to a wide range of poor socioeconomic outcomes in adulthood (Grantham-McGregor et al. 2007).

To reduce social inequality and promote early child development, preschool interventions are often considered to hold considerable promise. In fact, preschool programs for disadvantaged children such as Head Start, the Perry Preschool program, or the Carolina Abecedarian project, have shown high rates of return in the US (Ludwig et al. 2007; J. J. Heckman et al. 2010; J. L. García et al. 2017; J. Heckman et al. 2013) and also in Canada (Algan et al. 2022), Denmark (Rossin-Slater et al. 2020), Argentina (Berlinski et al. 2009), and Spain (Felfe et al. 2015). However, attending preschool does not always improve child development: several studies show that attending public preschool does not improve the average child's subsequent educational outcomes in Norway (Drange et al. 2016) France (Goux et al. 2010), Denmark (Gupta et al. 2010), Canada (Baker et al. 2008), and Cambodia (Bouguen et al. 2018). These heterogeneous impacts of preschool on child outcomes may be related to variations in the quality of preschool across contexts, but also to the substitution between preschool attendance and parenting at home. A large literature points to the crucial role of parental stimulation which has consistently large positive impacts on child development (see Britto et al. 2017 and Attanasio, Cattan, and Meghir 2021 for recent literature reviews). Attending preschool should thus be seen as choosing one mode of education instead of another, so the outcome depends on the relative productivity of preschool and parents in producing child development.

This paper examines the relative productivity of parents and preschools in child development in the context of concentrated Roma communities in Bulgaria. Bulgaria has one of the largest Roma populations in Europe, and the vast majority of Roma live in marginalized conditions: in 2014, 86% of Roma lived with an income below the national at-risk-of-poverty threshold, against 22% in the general population (EU-MIDIS 2017). With 66% of Roma children aged 4-6 participating in preschool compared to 89 percent of the ethnic Bulgarian children (EU-MIDIS 2017), encouraging the participation of Roma children in preschool seems a natural solution to reduce subsequent educational gaps with Bulgarian children. We use a 2 x 4 arm randomized control trial involving 5,772 children across 236 poor communities with large shares of Roma to test several combinations of information, removal of kindergarten fees, and financial incentives that aimed at increasing preschool participation of 3-6 year-old children, and measure the subsequent one-year impact on children' preschool participation, parental home stimulation, and child development outcomes.

The results show that the interventions substantially increased preschool participation of children who were not already registered in preschool at baseline: their enrollment rate was 26 percentage points higher and their attendance rate 0.40 standard deviations higher than in the control group. We find no consistent differential impacts across all combinations of the information awareness campaign, free access, and financial incentives. The interventions also induced a 0.25 standard deviation decrease in parental investment in educational activities with their child, which indicates that parents substitutes preschool for their own investment. Surprisingly, the interventions had a negative impact on the development of children not registered at baseline: numeracy scores are 0.14 standard deviations lower in the treatment group compared to the control group after one year (literacy and socioemotional skills scores are also 0.08 and 0.12 standard deviations lower than in the control group, but these differences are not statistically significant).

An important question is whether the reduction in parental investment is the cause of the reduction in child development, or if it is due to difficulties encountered by children from poor communities in the formal education system. To answer this, we estimate a structural model of a child development production function and quantify the degree of technical substitution/complementarity between school attendance and parental investments and their relative contribution to child development. Specifically, we specify a CES child development production function and use our experimental setting to estimate its parameters. The originality of our estimation is that the contribution of preschool and parents to child development can be identified using the eight randomized treatment conditions as instruments, which is rare in this literature. Specifically, we take a three stage approach. In a first stage we non-parametrically estimate the expected levels of school attendance and parental investment conditional treatment type indicators. Then, we estimate the production function taking a control function approach using the previously obtained conditional expectations as instrumenns.

The results show that the elasticity of substitution varies between 0.38 and 0.44 depending on the skills, which indicates that the degree of substitutability between parental investment as school attendance is high but statistically different from perfect substitution. Thus, an increase in preschool attendance that crowds out parental investment is associated with a decline in the marginal productivity of preschool, and *vice-versa*. By contrast, increasing both preschool attendance and parental involvement would increase each input's marginal productivity. This finding makes a strong case for policies favoring both access to formal education and parental investment in young children, which are rarely promoted together.

Secondly, the structural estimation shows that the contribution of parental investment to child development is as large as the contribution of preschool attendance. This result is particularly important given that most parents in are sample are Roma, poor, and low-educated. It means that even extremely disadvantaged parents have a large contribution to child development. At observed levels of input utilization, the marginal product is larger for parental investment than for preschool attendance whatever the ethnicity of the parents and their initial enrollment status. The higher parental marginal product holds for literacy, numeracy, and socioemotional skills, although not for motor skills. Our counterfactual calculations show that closing the gap in parental investment between minority and majority children would reduce by the observed differences in literacy, numeracy, and socioemotional skills between these children by 33, 58, 73 and 67 percent respectively whereas closing the gap in school attendance would reduce them by only 19, 27, 33 and 38 percent. The implication of our findings is that the best policy to foster disadvantaged children' development would combine preschool participation with a strong emphasis on parental involvement.

This paper makes three important contributions. First, we contribute to the literature on the demand for education. Our paper shows that eliminating the direct costs of preschool increases demand for preschool and reduces by half the gap between disadvantaged and advantaged children. Importantly, our paper shows how demand for preschool varies as a result of the interaction between price and information, at four price levels both positive and negative. The literature documents the effect of price (Baker et al. 2008; Duflo et al. 2015; Duflo et al. 2021; Giacobino et al. 2022; S. García et al. 2022; Lucas et al. 2012) and the effect of information (Jensen 2010) separately (on education participation at pre-primary, primary, and secondary), this paper provides first evidence that information and free access are largely substitutes and that negative prices (conditional cash transfers) do not further increase demand as soon as children are offered free access and parents are informed about preschool benefits.

Second, our paper contributes to the literature on child human capital production function. This literature has made recent progress that uncovers important interactions between—and dynamics of—children's health, cognitive skills, socioemotional skills, and parental investments (Attanasio, Cattan, Fitzsimons, et al. 2020; Attanasio, Bernal, et al. 2020; Cunha et al. 2010; Attanasio, Meghir, et al. 2020; Del Boca et al. 2014). Our paper provides the first structural estimation of a production function with preschool and parental investment as inputs. The finding that parents and preschool are complements in child human capital production function is an important addition to this literature because it points to the caveat of promoting preschool without considerations for parental investment. We provide the first structural estimation of parents' and preschool relative productivity, which complements nicely Özler et al. 2018 who provide experimental evidence that improvements in classroom quality does not translate into improvements in child outcomes unless it is combined with an increase in parental inputs. An important policy implication is that parental involvement and preschool participation should be promoted together.

Finally, our paper contributes to the vast literature on the importance of parents in early child development. A large consensus emerged in the literature regarding the large impacts of parenting programs (Britto et al. 2017; Attanasio, Cattan, and Meghir 2021). Our paper adds an important contribution to this literature by showing that disadvantaged parents, even when they are uneducated and when no support is provided to improve their parental skills, contribute greatly to child human capital formation. This finding goes along with recent work showing that disadvantaged parents underestimate the returns to their investment in child education, and that these beliefs are detrimental for actual investment decisions (Attanasio, Cunha, et al. 2019; Attanasio, Boneva, et al. 2019; Attanasio, Cattan, and Meghir 2021). As a result, disadvantaged parents should be considered as fundamental actors in child human capital formation, and be encouraged to intensify their involvement.

The rest of the paper is organized as follows. Section 2 presents the context of the study in Bulgaria. Section 3 presents the program design. Section 4 presents the methodology and data of the randomized controlled trial, and Section 5 presents the experimental results. Section 6 presents the methodology of the structural estimation of child human capital production function and its results. Finally, Section 7 concludes.

2 Context

The study takes place in poor communities in Bulgaria with large shares of Roma. There are approximately 10-12 million Roma in Europe, the largest ethnic minority in Europe. The living conditions among the majority of Europe's Roma minority are dire. A 2016 survey by the Fundamental Rights Agency[1] finds that, in nine European Member States, 80 percent of Roma are at risk of poverty compared with an EU average of 17 percent. The EU's largest ethnic minority faces intolerable discrimination and unequal access to vital services. Providing Roma children with an equal start in life compared with their non-Roma peers is essential to break the cycle of intergenerational transmission of poverty. However, the same survey finds that among Roma youth aged 18-24, the proportion with at most lower secondary education is between 57 and 92 percent across the nine countries, falling far short from attaining the EU2020 target of 10 percent for early school leaving. And, although improving in some countries, the education gap starts early: the same survey reports that 53

[1] https://fra.europa.eu/en/publication/2016/second-european-union-minorities-and-discrimination-survey-roma-selected-findings. The report is based on a large-scale survey that collected information on almost 34,000 persons living in Roma households in nine European Union (EU) Member States

percent of preschool aged Roma children attend early childhood education, often less than half the proportion of children their age from the general population, and well below the 95 percent EU target.

In 2011, the year with the most recent data on preschool enrollment rates for both Roma and non-Roma in Bulgaria before the 2014 start of this evaluation, 45% of Roma children were estimated to be enrolled in preschool compared with 75% as a national average (incl. Roma) (World Bank, 2012). Legal amendments from September 2010 provided for 2 years of compulsory preschool education to encompass all 5 and 6-year-olds. The municipalities had 2 years to ensure that they were able to accommodate all children. For 5 and 6-year-old children, the educational services are free of charge. However, kindergartens charge fees to cover additional non-staff costs such as school supplies, toilet paper, and required medical examinations. At the time this study began in September 2014, participating in some form of preschool program was thus de jure compulsory and free of educational service charge for the 5 and 6-year-old children. In 2014-2015, at the time of our study implementation, However, whether these legal provisions had been successful in eliminating the costs and raising preschool rates among Roma children was not clear by the start of our study.

Furthermore, at the time of our study, families in Bulgaria with monthly incomes less than or equal to BGN 350 per family member were entitled to receive a child allowance for each child aged 1 or more. In 2014, the monthly child allowance was BGN 35 per month. Following a legal amendment in 2013, for children aged 5 and 6, this allowance was made conditional on the child regularly attending a preparatory preschool group (either half-day or full day), unless this is impossible due to the condition of the child's health.

Kindergartens serve 3 to 6-year-old children (but only the 5-6 year old groups are referred to as preschool). Full-day kindergartens, as opposed to half-day, are the preferred modality for early childhood education in Bulgaria. In 2014, approx. 92% of all children enrolled in public early child education programs attended a full day kindergarten. Children study in the morning from 8 to 12, they are provided morning and afternoon snacks and lunch, then take a nap, and then play until 4 or 5pm. The cost of full-day kindergartens is shared between the central government, local governments and families, from which the municipalities collect the kindergarten fees. The central government funding covers pedagogical and part of the running non-wage operating costs and comes from per capita based block grants transferred to municipalities, which in turn are channeled in full to the full-day kindergartens using a formula, mostly according to the number of children enrolled in them. The average per child governmental financing of full-day kindergartens was BGN 1655 in 2014.

Local government funding covers infrastructure maintenance, as well as part of the running non-wage operating costs (food, transportation, materials and similar items) through the kindergarten fee revenues. The full-day kindergarten fees paid by families are determined by local governments within a centrally defined framework (the Local Taxes and Fees Act) and are collected from parents/primary caregivers directly by the kindergartens. Some municipalities set daily fees; others determine a monthly fee for full attendance that is prorated for the actual days attended; yet others apply a combination of the two methods. In a few cases, fees differ monthly according to the cost of the food. The average annual per child amount of full-day kindergartens fees was estimated BGN 176, down by 2% from BGN 183 in 2014, according to official data.

3 Program Design

The primary goal of the impact evaluation was to assess ways by which to increase kindergarten enrolment and attendance, particularly among vulnerably minority children. The program, called Springboard for School Readiness (SSR), was jointly designed with the World Bank (WB) and the Trust for Social Achievement (TSA) in Bulgaria. It was designed as a demand side set of interventions to test three main reasons why children may not enroll and regularly attend pre-school: first, parents may lack awareness around the benefits that kindergartens may provide around early stimulation of children, or awareness around more practical issues, such as enrollment procedures, fees, or the nature of the curriculum. Second, parents may not be able to afford the kindergarten attendance fees. Third, parents may be reluctant to enroll their children for other reasons. For example, out of concerns for discrimination, a preference to keep the child at home, etc. To address each of these three reasons, the program had the following components:

Free access to full-day kindergartens The intervention intended to cover the full cost of education to eligible households so as to ensure that affordability is not an obstacle. The cost of education includes attendance fees as well as other formal costs: school supplies for 3-4 year old children and laboratory tests (stool and urine) required upon enrollment or after prolonged absences. The financial contributions to kindergartens were directly paid to the kindergartens by the program partner. Since the 2010 legal provision making preschool compulsory for 5 and 6 year-old children also requires an absence of fees for the educational services (but not for care like food, toilet papers, etc.), the intervention is expected to especially benefit younger children aged 3-4, at least to the extent that the completely cost-free alternatives (half-day programs delivered at kindergartens and primary schools) are available only for children aged 5 and 6 years.

Free access to full-day kindergartens + small incentive This intervention offered the same as the previous one and added a small financial incentive: a monthly bonus was offered to the family on the condition that the child attended full-day kindergartens daily (with the exception of absences as a result of illness). The monthly bonus was kept small at BNG 7 (approximately Euro 3.5), which represents 20% of the child benefit allowance. It thus serves as a nudge more than a financial gain. If families tend to prefer that their child attend kindergarten but fail at doing so because of time inconsistencies and lack of attention, a small amount may be sufficient to increase kindergarten participation. Again, the impact on parental behaviour of this intervention is expected to be larger for 3 and 4-year-old children to the extent that this financial incentive is conditioned on attendance, unlike the child allowances for children aged 3 and 4 years, which is paid by the government unconditionally.

Free access to full-day kindergartens + large incentive This intervention offered the same as the previous one except that the incentive is larger: BNG 20 (approximately Euro 10), which represents 60% of the child benefit allowance. The larger incentive may be more efficient at increasing kindergarten participation than the small one in case parents have strong reasons why their kid does not attend kindergarten: they do not like kindergarten education, they have important opportunity costs of time going to kindergarten and back home, etc. Similarly to previous interventions, the impact of this intervention is expected to be larger for 3 and 4-year-old children, but considering the larger amount, impacts on parental behaviours for the 5-6 years olds cannot be ruled out.

Information The intervention consisted of five public community meetings with duration of 45 to 90 minutes each at the kindergartens organized by the 23 implementing NGOs, for a total of about 6 hours over 10 months. The information campaign did not deal with parental skills and appropriate home-learning environment. The intervention provided parents with information about the benefits of kindergarten education in an effort to raise awareness about its importance for the further education and professional life of their child, and to encourage parents to feel more at ease in the kindergarten environment, with session 2 being a presentation by the teachers and session 3 an "open house" were parents were invited to class. These also promoted the interaction between parents and kindergarten officials. The concept was thus a "mediation and information campaign." All households with children ages 3-6 were invited to join the meetings through advertising by social workers and posters. The design and content of the information sessions are detailed in Appendix X.

To implement the programme, TSA partnered with 23 Bulgarian NGOs and collaborated with local kindergarten and municipal authorities. All interventions were implemented in full-day kindergartens (hereafter, KG). 16 of the 23 contracted NGOs were Roma-led. TSA trained each of the NGOs and each of full-day kindergarten principals participating in the financial interventions.[2]

4 Methodology

4.1 Sampling procedure

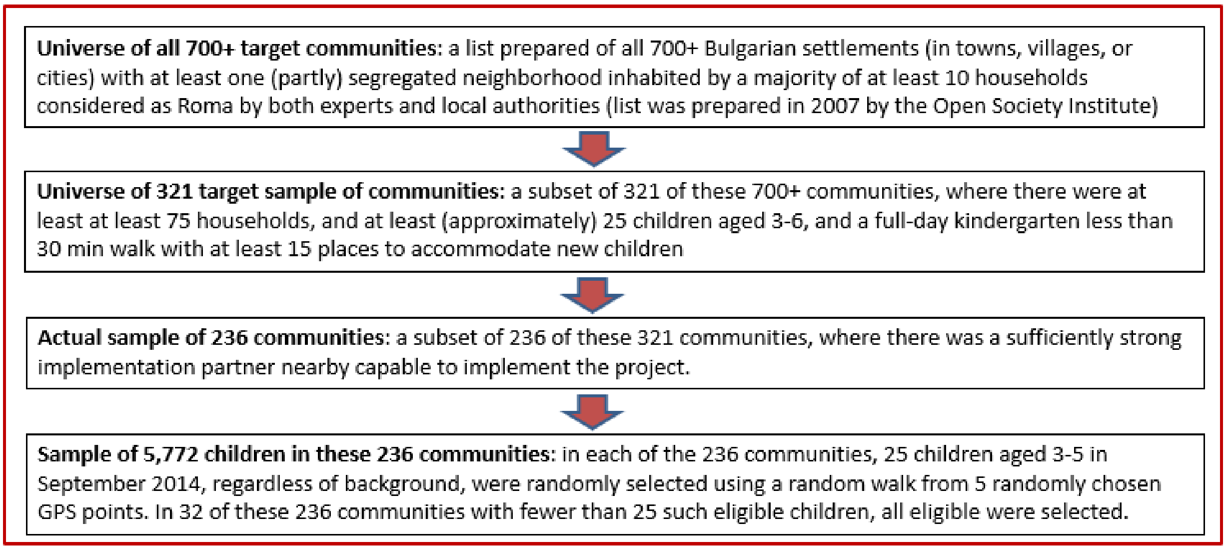
The intervention and the study targeted segregated Roma communities. In 2007, a community listing was prepared by the Open Society Institute (OSI) containing all Bulgarian settlements with at least 10 households considered as Roma by both experts and local au-

[2] An implementation constraint in some communities was that there was not enough room to enroll all the eligible children that would like to enroll. This is not a constraint for the evaluation as this is something that is part of the reality on the ground.

thorities. The list entailed more than 700 communities. The following criteria were used to select the communities for the evaluation: they should have least 75 households and 25 children age 3 to 6 by September 2014; the distance to full-day kindergarten should be less than a 30 min walk (unless free public transportation was provided); there should be availability for at least 15 additional children in the local full-day kindergarten. These criteria selected 321 communities of the 700 listed by OSI. From these, 236 communities were selected where there was a sufficiently strong local NGO capable to implement the project, and a preliminary agreement of the local mayor and full-day kindergarten principal to participate in the evaluation.

In April 2014, in each of these 236 communities, OSI identified 25 children who would be aged 3-5 at the start of the September 2014 school year, regardless of their background. If there were more than 25 children in a given community, OSI randomly selected 25 using a random walk from five randomly chosen GPS points. If there were fewer than 25 eligible children, all were selected. In other words, all 3-5 year-old children living in the community had an equal chance. All participant families provided informed consent before joining the study. Altogether 5,772 children were selected. Figure 1 summarizes the selection procedure:

Figure 1: Sample selection procedure



4.2 Data collection

A baseline survey was carried out in April-May 2014 to collect information on the communities, the kindergartens, and the households. In each community, the mayor, the kindergarten director, and a community leader were interviewed, as well as the 5,772 sampled households. The baseline survey captured information on demographics, education, and employment of all household members, including information on literacy and numeracy, as well as information on child rearing/interaction practices and attitudes toward parenting and education. It also included information on assets and dwelling characteristics. There were 427 refusals (7%).

The endline survey was conducted by OSI in April-May 2015 and covered 5,158 of the 5,772 baseline households, equivalent to an attrition rate of less than 11%. The endline survey covered most of the same baseline survey questions. In addition, the endline survey included direct measures of child development, using the International Development and Early Learning Assessment (hereafter, IDELA), developed by Save the Children. IDELA involves four developmental areas: motor, language and early literacy, math and problem solving, and socio-emotional development. The questionnaire constitutes a core of 24 small exercises that span these four developmental domains. The questionnaire was administered at home by a trained enumerator. The original questionnaire was written in Bulgarian, but enumerators adjusted to children's preferred language. For this purpose, enumerators were selected by OSI in order to fit local communities. Save the Children provided training to the enumerators.[3]

Finally, three unannounced kindergarten attendance checks were carried out by OSI during April-May 2015 covering all eligible children. During these checks, someone from the OSI team would make an unannounced visit to the kindergarten and verify which children

[3]In 2011, Save the Children completed a comprehensive review of the existing child development assessments and documented a number of important limitations with existing tools (targeting specific skill areas or a specific age group, reliant on parent or teacher report rather than directly assessing children's skills, expensive, or time-consuming). Most importantly, the majority of existing tools had been used primarily in high income countries, such as the United States, United Kingdom and Australia, making them difficult to adapt and easily use across countries with diverse populations and resource-poor settings. Save the Children conducted in depth quantitative analysis of item functioning, internal consistency, inter-rater reliability and construct validity. With testing and input from multiple country teams over three years, the tool has resulted in 2015 in a 22-item assessment that balances the three key dimensions discussed above: psychometric rigor, feasibility, and international applicability. As a result, IDELA is easily translated and administered in varied cultural contexts, and has strong reliability and validity. See Appendix X for examples of items included in the IDELA questionnaire used for this project.

were present and which children were not. In some cases only two unannounced visits were possible before the end of the school year.

4.3 Outcomes of interest

4.3.1 Preschool attendance

We consider two measures of school participation: enrollment and attendance. These measures were constructed combining data collected during the baseline and endline surveys as well as records from the unannounced visits to the schools and administrative data on enrollment provided by the bulgarian authorities.

Enrollment is a minimal notion of school participation; we measured it at the start and the end of the intervention. At baseline, enrollment is self-reported and comes from the answer to the question *Is the child registered in a Kindergarten or pre-school?*. At endline, we had access to administrative data from which we constructed an enrollment indicator. Naturally, there are some discrepancies between the self-reported and administrative data, In our final sample, there is no response to the enrollment question for 1.3% of the children, 3% declared not to be enrolled but were found in the school records and 7% reported to be enrolled but were not found in the administrative data. Despite this discrepancies, we found no substantial differences when using one or the other variable in our analyses.

The central objective of the intervention was to increase school participation; as noted above, enrollment is not an ideal measure since children might enroll but never attend a school. For this reason, we also use the records from the unannounced visits to schools. There were three visits during the intervention, thus, the number of times each child was found ranges from 0 to 3. This count of visits is the main measure of school participation throughout the study. To facilitate comparison with measures of parental investment, we divided the attendance count by its standard deviation.

4.3.2 Parental investment

We measure parental investment using questions from surveys related to the interactions between the child and the parents. Each question asked whether in the past three days the parents had performed certain activities with their children. The specific activities were reading books, telling stories, singing, naming objects, playing with toys and going outside; the answer to each question was binary. To aggregate all the answers in a single score we took the sum of all the replies as the outcome of interest and divided it by its standard deviation. This variable was constructed using the exact same procedure and variables (but measured at different points in time) at baseline and endline.

4.3.3 Child development

To measure developmental outcomes we used the IDELA tests that were administered at the end of the intervention. The IDELA test is specifically designed to measure literacy, numeray, socioemotional and motor skills and has questions specifically tailored for each ability. This provides a simple way to measure the four skills of interest by just calculating the score that each child obtained in the associated portion of the test. To facilitate comparison without ignoring potential differences in variation across abilities as well as the relative difficulty of each question we divided all the scores by their sample standard deviation.

4.3.4 Socioeconomic control variables

Throughout the analysis we include socio-demographic covariates in order to control for potential confounding factors such as gender and age. Furthermore, we explore potential determinants of parental investment and school participation such as discrimination, parental expectations and the educational attainment of parents. These variables are all constructed from the baseline survey and remain fixed throughout the study to guarantee that they are uncorrelated with treatment assignment.

Age and gender are self reported and taken directly from the baseline survey. Parental expectations is recovered from the question *What is the maximum level of education that you expect your child to obtain?* however, most of the variation concentrates between parents that expect their child to complete secondary education or not, thus, for most of our analyses we just include a variable that indicates the expectations of parents regarding secondary education. Similarly, for parental education we just distinguish between parents who completed secondary education and those who not.

Since budget constrains are likely to affect parental investment and school attendance decisions, we also include them as relevant controls in our analysis. We use a measure of self reported income as well as an index constructed from the answers to a series of questions about amenities and appliances owned by the household members.

Finally, in our setting discrimination towards minority children is likely to undermine school participation for Roma and Turkish children. This could happen through several channels, for instance, the children could be discriminated by their classmates or even by the teachers at school. Furthermore, minority parents might also be deterred from enrolling their children if they perceive that they are not fully accepted by their majority peers. Aside from being a potential confounder in our analysis, if discrimination is harming the opportunity for minorities from such early stages, detecting and quantifying its effect becomes a matter of central relevance for researchers and policy makers.

To account for discrimination and measure its effects, we construct an index for each household aimed at capturing discrimninating practices that are specifically targeted against Roma and Turkish children and their parents. The index is the sum of three questions that ask the parents about their beliefs that the children will be maltreated at school by their peers, the teachers and the parents of other children. There is a concern that parents who care more about education are more likely to enroll their children and actively participate of school activities which in turn makes them more likely to notice discrimination. For this reason, the index for each household only includes the replies from minority parents from other households that belong to the same community.

4.4 Experimental Design

A public lottery to assign the interventions was organized on June 30, 2014. Each of the 236 mayors and kindergarten principal had provided consent to participating. The lottery was held in the presence of representatives of the Ministry of Education, the National Association of Municipalities in Bulgaria, mayors, directors of full-day kindergartens, NGOs, and others. Using information from the 2014 baseline survey, the random assignment was stratified by the implementing NGO and the number of available spots in the full-day kindergarten. Each stratum contains eight communities.

Within each stratum, we implemented a two-step randomization. First, two communities were randomly assigned to one of four financial conditions: free kindergarten, free kindergarten + BGN 7 bonus conditional on attendance, free kindergarten + BGN 20 bonus conditional on attendance, and no financial intervention. Second, within each pair of communities assigned to a financial condition, one community was assigned to the information campaign and not the other one. This resulted in eight experimental group of 29 or 30 communities each, as represented in Figure 2:

Figure 2: Experimental design

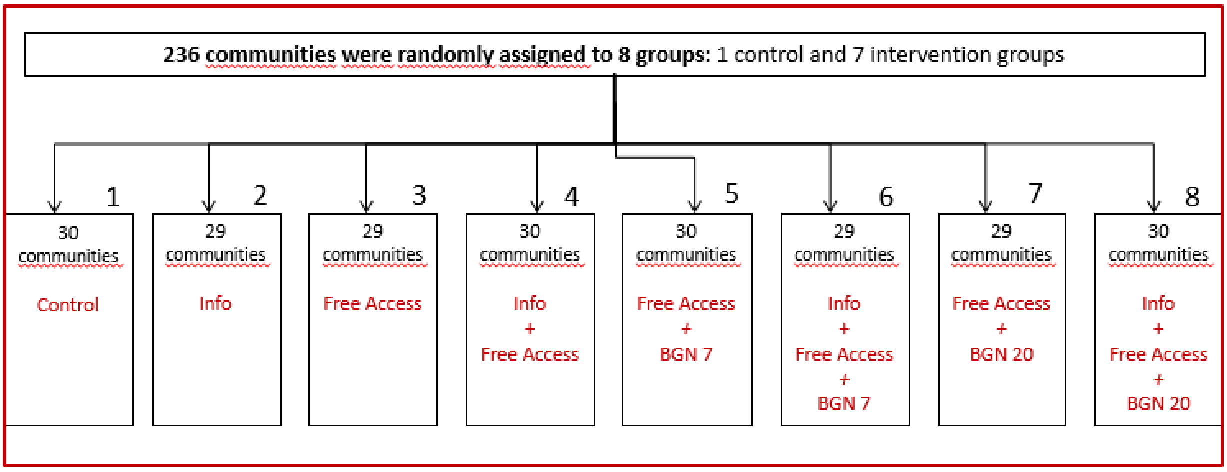


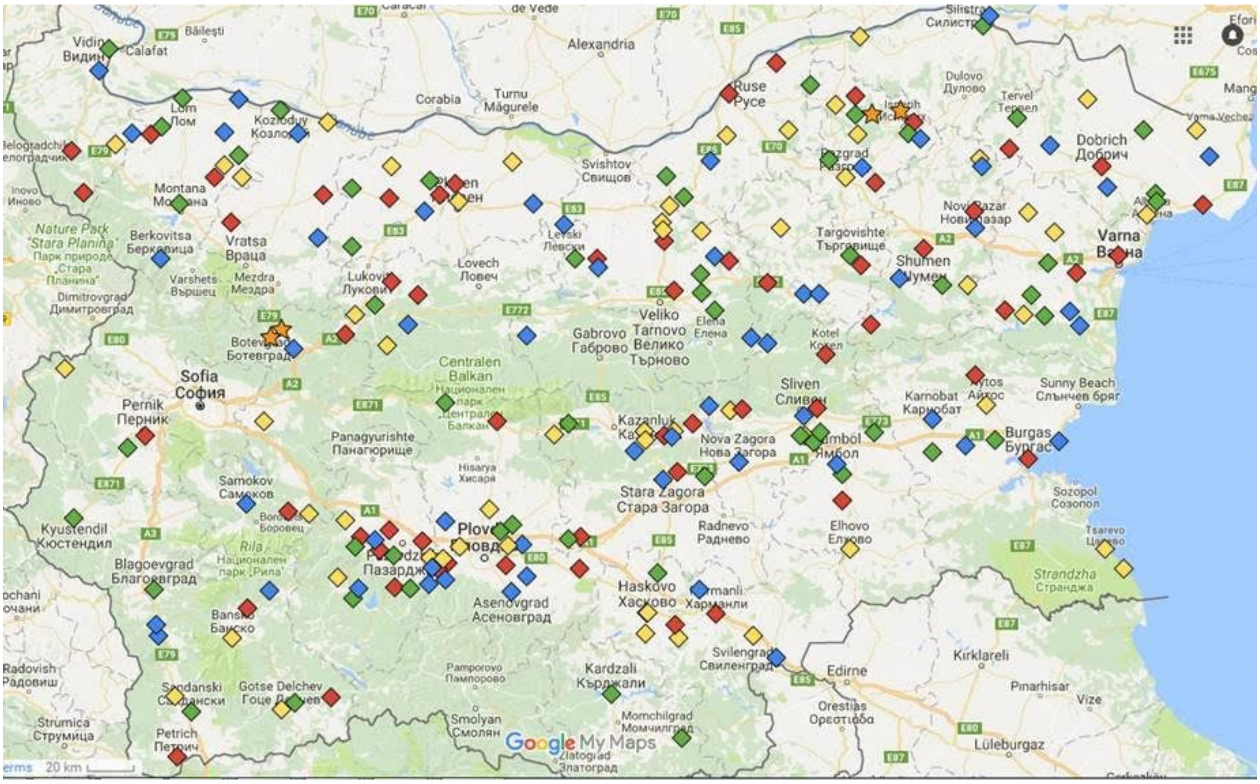
Figure 3 shows the geographical distribution of the financial conditions across communities on a map of Bulgaria. It does not represent the allocation of the information campaign across the communities, which was equally distributed to half of the communities within each financial intervention. One color stands for one financial intervention: yellow for control communities, blue for free access communities, green for free access + BGN 7 incentive communities, and red for free access + BGN 20 incentive communities. Four communities in which TSA carried out a pilot study during 2013-2014 (and for that reason are excluded from this evaluation) are displayed as orange stars.

5 Results

5.1 Sample selection

The study covered 5,158 children, however, the endline survey was not administered to all the children who started in the study. For this reason, our final sample comprises 4,839 children

Figure 3: Geographical distribution of the 236 communities, by financial condition



of which 3,764 are of Roma or Turkish ethnicity. The final sample corresponds to 93% of the baseline sample.

In addition to attrition, there are missing values or record mistakes for some of the responses to the baseline survey. Instead of dropping these observations we used a multiple imputation by chained equations algorithm. The method consists of assuming missing observations take the median value, then regressing all observations on the remaining observables and finally, replace the median imputations by the predictions that the resulting model produces.

5.2 Baseline sociodemographic and developmental differences

We start by documenting the disadvantages that Roma and Turkish children face with respect to their majority peers. Despite having similar demographic compositions; majority children live in richer, more educated households, and their parents have higher expectations regarding their educational attainment. Naturally, this is correlated with higher parental investment as well as higher enrollment rates which translates into better performance in tests of cognitive and non-cognitive skills. Our results not only show the severe disadvantage that minority children suffer but also illustrate the multiplicity of dimensions in which it is realized as well as its potentially self-reinforcing nature.

Table 2 was obtained calculating the uncontrolled mean outcomes for minority and majority children and jointly testing for statistically significant differences across outcomes. To account for the multiplicity of hypotheses being tested we implement the bootstrap based method of Romano et al. (2005) using the Stata 16 command rwolf (see Clarke et al. 2020) The first panel shows the means of the literacy, numeracy, socioemotional and motor skills index at endline; we find significant differences in all four outcomes. The second panel repeats the previous exercise but including the whole sample and testing for baseline differences in sociodemographic characteristics as well as parental investment and school participation. We find a significantly lower enrollment rate for minority children as well as consistently lower levels of parental involvement across all the surveyed activities. This last result combined with the differences in cognitive and non-cognitive skills suggests the urgency of interventions aimed at increasing school participation and eliminating the barriers that prevent minority parents from investing in their children.

One plausible hypothesis is that minority parents are less informed about the gains from human capital investment or reluctant due to cultural reasons and historical circumstances. While we cannot directly test this hypothesis, we document in the second panel of 2 that, on average, Roma and Turkish parents have lower educational attainment and have lower expectations regarding the educational accomplishment of their offspring. This might explain the observed input differences which in turn translate into severe developmental disadvantages.

In addition to the differences in developmental inputs and outputs, table 2 also documents severe differences in economic status as proxied by income and the assets and dwelling index. Furthermore, we find that minority households perceive more discrimination toward them while not necessarily living in communities where discrimination is higher. This result is not surprising since not being the victim makes ethnic discrimination less noticeable. Yet, it highlights the need to bring awareness about the discrimination and lack of opportunity that minorities face.

Having shown the disparities between minority and majority children and how they extend beyond the scope of cognitive and non-cognitive abilities; we investigate to what extent sociodemographic and parental characteristics account for the differences in both developmental inputs. We do this using a linear regression model.

Formally, we we estimate the following equation:

$$y_{icpga} = \mu_g + \mu_a + \mu_{rt} + \beta X_{ihc} + \epsilon_i \tag{1}$$

In the previous equation, y_{icpga} represents the variable of interest for child i in community c whose gender and age are g and a respectively. The term μ_{rt} is a minority fixed effect while μ_g and μ_a are gender and age fixed effexts respectively. The matrix, X_{ihc} is comprised by the parental characteristics of interest (expectations regarding secondary education of the child and whether if the household head completed secondary education) along with a community-level index that measures percieved discrimination toward Roma and Turkish children at school and its interaction with a Roma and Turkish indicator. The term ϵ_i captures the remaining unobserved variation. Note that in the previoys analysis, there is no guarantee that the unobserved term ϵ_i is uncorrelated with the regressors. For this reason our results merely shed light on the correlation patterns but should not be interpreted as causal relations.

Table 3 shows the results from estimating multiple versions of equation 1 by ordinary least squares. We find that Roma and Turkish children are 17 percent less likely to be enrolled and (as column 2 shows) this difference is not attributable to demographics. Columns 3 and 4 add log of income and an indicator of parents expecting their child to complete secondary. We find that both income and parental expectations are highly correlated with enrollment and explain about 7 percent of the observed difference. In column 5 we add parental education to our specification but find that conditional on income and parental expectations parental education is not significantly associated with enrollment. Finally, we add discrimination to our analysis. Column 6 shows that the observed differences in enrollment are highly correlated with perceived discrimination. This suggest an explanation for the lower school participation of minority children and highlights the urgency of policies promoting the inclusion of minority children at school.

The seventh and eighth rows of table 3 show that minority children receive 0.31 less standard deviations of parental investment and that this difference is not caused by demographics. As in the case of school enrollment, we find that household income and parental expectations partially explain this gap (see columns 9 and 10) as well as parental education proxied by an indicator of the household head completing secondary. Finally, in column 12 we find that the interaction between perceived discrimination and the minority indicator is highly significant. This suggests that those children that are discriminated against are also the ones that receive less parental investment undermining their early development which might have severe long-lasting consequences.

Table 1: Variable measurement and construction

Variable	Data source	Type	Construction
Developmental skills			
Literacy	IDELA test	Z-score	Standardized average of the answers to the IDELA questions aimed at measuring literacy
Numeracy	IDELA test	Z-score	Standardized average of the answers to the IDELA questions aimed at measuring numeracy
Socioemotional	IDELA test	Z-score	Standardized average of the answers to the IDELA questions aimed at measuring socioemotional skills
Motor	IDELA test	Z-score	Standardized average of the answers to the IDELA questions aimed at measuring motor skills
School attendance			
Baseline enrollment	Baseline survey	Binary	Indicator of self reported enrollment
Enrollment	Kindergarden records	Binary	Indicator of appeareance in records
Attendance	Three unnanounced KG visits	Integer	Count of times the child was present
Parental investment			
Baseline investment	Baseline survey	Z-score	z-score of indicators of the parents spending time reading, playing, naming, singing, telling stories and drawing with their kids
Endline investment	Baseline survey	Z-score	z-score of indicators of the parents spending time reading, playing, naming, singing, telling stories and drawing with their kids
Demographics			
Sex	Baseline survey	Binary	Self reported
Age	Baseline survey	Binary	Self reported
Roma or Turkish	Baseline survey	Binary	Self reported (determined by surveyor when missing)
Parental characteristics			
H.H. completed secondary	Baseline survey	Binary	Self reported
Parents expect secondary completion	Baseline survey	Binary	Self reported
Discrimination			
Community perception	Baseline survey	Community avg. of Z-scores	Average of household's z-score comprised by indivators of the parents percieving discrimination from teachers and other parents

Table 2: Differences between Bulgarian and minority households

	Majority	Roma or Turkish	H0: Majority=Roma or Turkish F-statistic	RW p-value
Panel A: Control group differences in developmental outcomes				
Literacy	1.14	0.85	13.53	0.01
Numeracy	1.67	1.44	9.82	0.01
Socioemocional skills	1.47	1.33	4.85	0.01
Motor skills	1.12	0.9	6.82	0.01
Observations per ethnicity	148	483		
Observations	631			

	Majority	Roma or Turkish	H0: Majority=Roma or Turkish F-statistic	RW p-value
Panel B: Baseline differences in enrollment, parental investment and household characteristics				
Kindergarden enrollment				
Enrolled at baseline	0.78	0.61	37.8	0.01
Parental investment				
Read books	0.71	0.41	119.63	0.01
Told story	0.79	0.62	38.9	0.01
Sang	0.78	0.71	6.04	0.08
Went out	0.95	0.91	10.45	0.02
Played	0.79	0.74	3.56	0.27
Draw	0.78	0.64	24.67	0.01
Summary Index	1.77	1.46	37.39	0.01
Demographics				
Female	0.48	0.48	0.06	0.97
Age	4.76	4.75	0.09	0.97
Household characteristics				
H. Head Unemployed	0.3	0.42	18.41	0.01
H. Size	4.49	4.69	6.64	0.08
Income	6.02	5.61	59.49	0.01
Parental education				
Completed Kindergarden	1.2	1.37	183.04	0.01
Completed Secondary	0.48	0.12	157.78	0.01
Parental aspirations				
Age to stop education	21.73	19.31	116.53	0.01
Expects child to complete secondary	0.9	0.76	61.85	0.01
Percieved discrimination				
Household percieves Disc. Toward RT childern at school (Index)	-1.69	-1.65	6.37	0.08
RT Community percieves Disc. Toward RT childern at school (Index)	0.52	0.56	0.84	0.75
Observations per ethnicity	1,101	3,762		
Observations	4,863			

Notes: This table reports observed differences between Bulgarian (Majority) and Roma or Turkish children. The first two columns report averages for each population group and the last two report F-statistics and p-values for the null hypothesis that the observed differences are attributable to chance accounting for the multiplicity of hypotheses being tested using the method of Romano et al. 2005. The developmental outcomes are measured as the z-scores of different measurements that correlate with literacy, numeracy, socioemotional and motor skills. Demographics, household characteristics, school attendance , parental involvement, parental education and discrimination are calculated using the baseline survey. Baseline enrollment is self reported, The summary index of involvement is calculated as the z-score of all the reported involvement variables.

Table 3: Socio-demographic characteristics and baseline developmental inputs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent variable	Baseline enrollment						Parental investment					
Roma or Turkish	-0.17***	-0.17***	-0.15***	-0.10***	-0.09***	0.10***	-0.31***	-0.31***	-0.28***	-0.18***	-0.16***	-0.01
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.10)
log income			0.05***	0.03***	0.03***	0.03***			0.08***	0.06***	0.06***	0.05***
			(0.01)	(0.01)	(0.01)	(0.01)			(0.02)	(0.02)	(0.02)	(0.02)
Parents expect secondary completion				0.17***	0.16***	0.11***				0.27***	0.24***	0.25***
				(0.02)	(0.02)	(0.02)				(0.04)	(0.04)	(0.04)
Household h. completed secondary					0.04	0.02					0.25***	0.24***
					(0.03)	(0.02)					(0.05)	(0.05)
RT com. perceives disc. at school						-0.31***						0.14
						(0.04)						(0.09)
RT com. perceives disc. at school#RT						-0.20***						-0.19**
						(0.04)						(0.10)
Constant	0.78***	0.53***	0.25***	0.23***	0.20**	0.49***	1.77***	1.75***	1.24***	1.21***	1.04***	0.94***
	(0.02)	(0.04)	(0.09)	(0.09)	(0.09)	(0.06)	(0.05)	(0.07)	(0.14)	(0.14)	(0.14)	(0.17)
Age and Gender FE	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Mean of the dependent variable	0.649	0.649	0.649	0.649	0.649	0.649	1.532	1.532	1.532	1.532	1.532	1.532
Observations	4,863	4,863	4,863	4,863	4,863	4,863	4,863	4,863	4,863	4,863	4,863	4,863
R-squared	4,863	4,863	4,863	4,863	4,863	4,863	4,863	4,863	4,863	4,863	4,863	4,863

Notes: This table reports observed correlations between baseline developmental inputs (school enrollment and parental investment) and observable characteristics. The first and seventh columns report the uncontrolled difference, the second and eighth add age and gender indicators as controls, the third and ninth columns include log of income, columns four and ten add an indicator for the parents expecting their child to complete secondary, columns five and eleven add an indicator of the household head completing secondary, finally, columns six and twelve add an index of discimination at the community level and its interaction with a minority indicator. Standard errors are clustered at the community level.

6 Intervention design

- Balance tables

Figure 4: Distribution of treatment arms across selected communities

Intervention	No financial intervention	Free access to preschool	Free access + 7 lev conditional on attendance	Free access + 20 lev conditional on attendance
No meeting	30 communities	29 communities	30 communities	29 communities
Information and mediation meetings	29 communities	30 communities	29 communities	30 communities

Table 4: Table of balance p-values

	Information	Free access	Free acces+Info.	7 lev	7 lev+Info	20 lev	20 lev+Info
Developmental outcomes							
Literacy	1.000	1.000	0.980	1.000	1.000	0.990	0.990
Numeracy	0.663	0.248	0.455	1.000	0.980	0.525	0.475
Socioemocional skills	1.000	1.000	0.960	1.000	1.000	1.000	0.950
Motor skills	1.000	1.000	1.000	1.000	0.960	1.000	0.762
Kindergarden enrollment							
Registered at baseline	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Parental investment							
Read books	0.901	1.000	0.941	1.000	1.000	0.901	1.000
Told story	0.851	0.970	1.000	1.000	0.990	0.743	0.901
Sang	0.990	1.000	1.000	1.000	1.000	1.000	1.000
Went out	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Played	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Draw	1.000	1.000	0.980	0.950	1.000	1.000	1.000
Summary Index	0.980	1.000	1.000	1.000	1.000	0.990	1.000
Demographics							
Female	0.911	0.941	1.000	1.000	0.990	0.950	0.901
Age	1.000	1.000	0.891	1.000	0.980	0.990	1.000
Household characteristics							
H. Head Unemployed	0.683	1.000	0.980	1.000	1.000	1.000	1.000
H. Size	0.525	1.000	0.861	0.950	0.980	1.000	0.455
Income	0.525	1.000	0.614	0.980	0.990	0.475	0.990
Assets & Dewll. Ind.	0.990	1.000	1.000	1.000	1.000	0.950	1.000
Parental education							
Completed Kindergarden	0.911	1.000	0.950	1.000	1.000	1.000	1.000
Completed Secondary	1.000	1.000	0.564	0.941	1.000	0.990	0.802
Parental aspirations							
Age to stop education	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Expects child to complete secondary	0.832	1.000	0.970	1.000	0.752	1.000	0.901
Percieved discrimination							
Community percieves Disc. Toward RT childern at school (Index)	1.000	0.911	0.713	0.376	0.752	0.990	0.921
Observations	4,817	4,817	4,817	4,817	4,817	4,817	4,817

Notes: This table reports Romano-Wolf p-values obtained from 100 boostrap replications for the null hypothesis that treatment arm means are equal to control group means. Each column correspond to an invervention arm and each row to a different baseline variable.

7 Treatment effects

We measure the effect of the intervention on two dimensions; one that looks at inputs of the human capital accumulation process and measures the direct increase in enrollment and attendance along with the potential adjustments induced on parental investment. Second, we measure the effect of the intervention on cognitive and non-cognitive skills.

7.1 Estimation

The experimental design of the intervention would allow us to estimate the average treatment effects just by comparing outcomes in the control group and the different treatment types. However, there is a concern that attrition is non-random or that the effect of the intervention is confounded by socio-demographic characteristics. For this reason, we account for age, gender and parental characteristics such as expectations and educational attainment in our evaluation of the treatment effects. We do this under using machine learning techniques to approximate an additively separable, potentially non-linear nuisance function following the approach developed by Chernozhukov et al. (2018) commonly known as double-debasing.

Under the double-debasing approach, the outcomes of interest and treatment take-up are defined by the following two equations:

$$y_{icpga} = D_{it}\theta_t + g_0(X, a, g, rt) + \varepsilon_i, \ . \tag{2}$$

$$D_{it} = m_0(X, a, g, rt) + \nu_i, \ , \tag{3}$$

Where D_{it} is the treatment assigned to child i and $g_0()$ and m_0 are nuisance functions that we allow to depend on observable characteristics such as age, gender, ethnicity, parental expectations and educational attainment and discrimination percieved by the community as well as missing value indicators for all the variables included in our analysis. The terms ϵ_i and ν_i are assumed to have mean zero conditional on the treatment assignment and the components of the nuisance functions. As in the linear specification, D_{it} might just indicate the non-control observations or each of the eight different treatments separately.

To estimate the partially linear model described by equations 2 and 3 we use the double-debiasing machine learning algorithm of Chernozhukov et al(2018) as implemented by the Stata 16 package *pdslasso* developed by Ahrens et al (2018).

Naturally, the intervention is more likely to affect children that were not enrolled. For this reason, we perform both our linear regression and double-debiasing analyses on two different samples: one comprised by children that were not enrolled in kindergarten at the time of the baseline survey and another one comprised by those who were already enrolled.

Since the randomization varies across communities rather than households, we cannot rule out potential unobserved correlation between outcomes for children within the same community. For this reason, all the reported standard errors and inference results were obtained using cluster robust procedures at the community level.

7.2 Results

We find that the intervention was effective at increasing both enrollment and attendance but at the cost of a reduction in parental investment. Regarding cognitive and non-cognitive skills we find that the intervention had small or zero overall effects which suggest that the reduction in parental investment was just enough to offset the potential gains from increased school participation.

Tables 5 presents estimation results for the average treatment effects of the intervention on developmental inputs. The first panel shows the *pooled* effect, this is, the treatment effect that results from comparing the control group to all the treated observations regardless of their treatment type. Columns 1 and 2 show the effect on enrollment while columns 3 and four show the effect on the standardized number of unannounced visits in which the children was present. We find that there was a significant increase of 0.3 and 0.12 standard deviations on enrollment and attendance respectively among children that were not enrolled at baseline; along with 0.07 standard deviations higher enrollment rate among treated subjects after the intervention which suggest that it was also effective at reducing drop-out . Parental investment on the other was reduced as a response to the intervention. As columns 5 and 6 show, parental investment among children that were not enrolled at baseline decrease by 0.71 standard deviations.

The second panel of table 5 reports the average effect of each specific treatment arm.

We find a consistent increase in enrollment across all treatment types as well as significant increases in attendance among all but two treatment arms. Additionally, we find that the reduction in parental investment concentrated in the free access treatments as well as the ones that included 7 lev and 20 lev plus information.

Finally, table 6 reports estimates for the effect of the intervention on cognitive and non-cognitive skills. Here. we find not significant effects of the pooled treatment except for a marginally significant reduction in numeracy among children that were not enrolled at baseline, and a significant increase in motor skills among children that were already enrolled. The overall results suggest that the increase in enrollment and attendance was not sufficient to compensate the effects of the associated reduction in parental investment. Desegregating the effects across treatment arms reveals a similar pattern, we find some marginally significant reductions in numeracy and literacy ranging between 0.16 and 0.17 standard deviations as well as some increases in motor skills among children that were already enrolled that range from 0.17 to 0.02 standard deviations.

Table 5: Double debiased LASSO treatment effects on school participation and parental investment

Panel A: Pooled treatment effect

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Enrolled at KG		Attendance count		Parental investment	
Enrolled at baseline	Yes	No	Yes	No	Yes	No
Treated	0.04	0.26***	0.06	0.40***	-0.09	-0.25**
	(0.03)	(0.07)	(0.12)	(0.14)	(0.08)	(0.12)
Control group mean of dep. var.	0.91	0.60	1.85	0.99	1.49	1.27
Observations	3,156	1,707	3,156	1,707	3,156	1,707

Panel B: Individual treatment arm effects

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Enrolled at KG		Attendance count		Parental investment	
Enrolled at baseline	Yes	No	Yes	No	Yes	No
Information	0.04	0.24***	-0.22	0.32*	-0.01	-0.20
	(0.04)	(0.09)	(0.18)	(0.18)	(0.12)	(0.17)
Free access	0.07**	0.22**	0.15	0.45**	-0.18	-0.46***
	(0.04)	(0.09)	(0.15)	(0.21)	(0.12)	(0.13)
Fre access+ Info.	0.05	0.31***	0.08	0.42**	-0.07	-0.31**
	(0.04)	(0.10)	(0.14)	(0.18)	(0.12)	(0.16)
7 lev	0.07*	0.34***	0.10	0.57***	-0.05	-0.28**
	(0.04)	(0.09)	(0.19)	(0.18)	(0.10)	(0.14)
7 lev + Info	0.03	0.21**	0.04	0.25	-0.08	-0.08
	(0.04)	(0.08)	(0.14)	(0.17)	(0.11)	(0.18)
20 lev	0.02	0.25***	0.09	0.43**	-0.13	-0.17
	(0.04)	(0.08)	(0.15)	(0.19)	(0.11)	(0.15)
20 lev + Info	0.04	0.22**	0.12	0.37*	-0.08	-0.32**
	(0.04)	(0.09)	(0.15)	(0.21)	(0.12)	(0.15)
Control group mean of dep. var.	0.91	0.60	1.85	0.99	1.49	1.27
Observations	3,156	1,707	3,156	1,707	3,156	1,707
Equality of all treatment arms						
F-statistic	5.46	3.60	4.98	3.89	2.30	9.96
p-value	0.49	0.73	0.55	0.69	0.89	0.13

Notes: This table reports average treatment effects on developmental outcomes. The outputs are measured as the average z-score among a number of measurements correlated with literacy, numeracy, socioemotional and motor skills. The treatment effects were calculated using the Double Machine Learning method for partially linear models developed by Chernozhukov et al (2016). The method consists on using a machine learning method to estimate nuisance functions for both treatment adoption and outcomes of interest. The learner method is LASSO regression. The experiment consisted of 7 treatment arms plus a control group. The interventions where of two types, subsidies to school attendance and provision of information. The subsidies where free access and free access plus monetary transfers of 7 or 20 levs. Each subsidy group was divided in two subgroups, information and no information. Panel A reports coefficients pooling all treatment interventions. Panel B reports the average effect across all individual treatment arms.

Table 6: Double debiased LASSO treatment effects on cognitive and non-cognitive skills

Panel A: Pooled treatment effect

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable	Literacy		Numeracy		Socioemotional skills		Motor skills	
Enrolled at baseline	Yes	No	Yes	No	Yes	No	Yes	No
Treated	0.02	-0.08	-0.06	-0.14**	0.02	-0.12	0.13*	-0.01
	(0.05)	(0.08)	(0.04)	(0.07)	(0.05)	(0.08)	(0.07)	(0.08)
Control group mean of dep. var.	1.01	0.65	1.53	1.12	1.42	1.13	1.13	0.80
Observations	3,156	1,707	3,156	1,707	3,156	1,707	3,156	1,707

Panel B: Individual treatment arm effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable	Literacy		Numeracy		Socioemotional skills		Motor skills	
Enrolled at baseline	Yes	No	Yes	No	Yes	No	Yes	No
Information	0.00	0.05	-0.11	-0.08	-0.03	-0.04	0.08	0.03
	(0.07)	(0.09)	(0.07)	(0.08)	(0.07)	(0.09)	(0.09)	(0.09)
Free access	-0.04	-0.11	-0.18***	-0.16*	-0.05	-0.07	0.11	-0.03
	(0.06)	(0.10)	(0.06)	(0.08)	(0.07)	(0.09)	(0.08)	(0.11)
Fre access+ Info.	0.05	-0.13	-0.05	-0.17*	0.05	-0.19**	0.17*	-0.08
	(0.06)	(0.11)	(0.06)	(0.09)	(0.07)	(0.10)	(0.09)	(0.12)
7 lev	0.05	-0.12	0.06	-0.15	0.07	-0.18*	0.11	0.00
	(0.07)	(0.09)	(0.06)	(0.10)	(0.08)	(0.10)	(0.09)	(0.10)
7 lev + Info	0.06	-0.08	-0.03	-0.09	0.05	-0.12	0.16*	0.03
	(0.07)	(0.10)	(0.05)	(0.08)	(0.06)	(0.09)	(0.10)	(0.10)
20 lev	0.03	-0.05	-0.02	-0.14*	0.05	-0.06	0.10	-0.02
	(0.06)	(0.10)	(0.05)	(0.08)	(0.06)	(0.10)	(0.09)	(0.10)
20 lev + Info	0.01	-0.08	-0.06	-0.14	0.01	-0.17	0.20**	-0.02
	(0.06)	(0.11)	(0.05)	(0.09)	(0.06)	(0.10)	(0.09)	(0.11)
Control group mean of dep. var.	1.01	0.65	1.53	1.12	1.42	1.13	1.13	0.80
Observations	3,156	1,707	3,156	1,707	3,156	1,707	3,156	1,707
Equality of all treatment arms								
F-statistic	3.50	8.20	14.3	2.65	5.03	7.21	4	1.58
p-value	0.74	0.22	0.026	0.85	0.54	0.30	0.68	0.95

Notes: This table reports average treatment effects on developmental outcomes. The outputs are measured as the average z-score among a number of measurements correlated with literacy, numeracy, socioemotional and motor skills. The treatment effects were calculated using OLS controlling for gender, ethnicity and age of children. The experiment consisted of 7 treatment arms plus a control group. The interventions where of two types, subsidies to school attendance and provision of information. The subsidies where free access and free access plus monetary transfers of 7 or 20 levs. Each subsidy group was divided in two subgroups, information and no information. Panel A reports coefficients pooling all treatment interventions. Panel B reports the average effect across all individual treatment arms. The experiment consisted of 7 treatment arms plus a control group. The interventions where of two types, subsidies to school attendance and provision of information. The subsidies where free access and free access plus monetary transfers of 7 or 20 levs. Each subsidy group was divided in two subgroups, information and no information. Panel A reports coefficients pooling all treatment interventions. Panel B reports the average effect across all individual treatment arms.

8 The production function of developmental skills

We estimate a production function of developmental skills that takes parental investment and school attendance as inputs. This allows us to calculate the marginal product of both inputs under different counterfactual scenarios as well as the degree of complementarity that exist between them. Moreover, estimating a production function allows us to explore if some children benefit more from increases in school attendance and how this correlates with parental characteristics. Our findings rationalize the results of the intervention and show that an effective policy must take into account the substitutability between inputs and the adjustment decisions made by parents when their relative cost changes.

8.1 Substitution patterns between parental investment and school attendance

To asses the marginal effect of both inputs and understand their interaction we require to specify a tractable production function that does not impose strong restrictions on the substitution rate between parental investment and school attendance. The CES function is appealing for two main reasons. One is its functional form that can be estimated using control function techniques without the need to instrument for an interaction term. The second is that it does not impose strong restrictions on the degree of substitutability between inputs, even allowing for perfect substitutes and complements in limit cases.

In addition to school attendance and parental investment, there are other factors such as the skills and effort of parents, or children characteristics like age and gender that determine the realized developmental skills. Some of these determinants of productivity might be unobserved by the econometrician. Furthermore, it could be that even the parents do not observe or consider all these determinants at the time they choose attendance and parental investment for their children.

The previous discussion motivates the following empirical specification:

$$y^d_{igap} = exp\big(\gamma^d X_{ip} + \eta^d_{ip}\big)\big(\beta^d I^{d^\rho}_{ip} + (1-\beta^d)S^{d^\rho}_{ip}\big)^{\frac{1}{\rho^d}} \tag{4}$$

Here $d \in \{Numeracy, Literacy, Motor, Socioemotional\}$ denotes the developmental skill of interest. The outcome for child i whose gender is g, is a years old and has parents p is y_{igap}. The first term represents parental productivity which is determined by parent and child observable characteristics (X_{ip}) such as the education and expectations of parents and baseline enrolment plus an unobserved component η^d_{ip} that comprises other relevant variables, some of which might be unobserved by parents.

Parental investment (I) and school attendance (S) enter through the second term. The parameter β^d determines how intensive developmental skills are in each of the two inputs. Values of β^d close to one imply a larger contribution of parental investment while values close to zero represent a production function that is intensive in school attendance. We assume constant returns to scale (*i.e.* $\beta^d \in [0,1]$) The degree of complementarity between both inputs in the production of d is dictated by the parameter ρ^d which takes values between $-\infty$ and 1. As ρ^d tends to infinity both inputs become stronger complements while as ρ^d tends to 1 the production function tends toward perfect substitution between inputs. For reference, note that as ρ approaches 0 the functional form resembles a Cobb-Douglas production function.

8.2 Endogeneity and unobserved productivity

The main concern for identification is the endogeneity of school attendance and parental investment with respect to the unobserved productivity components. For instance, parents could decide to invest more if their children performs well at school. Or, in the opposite case, parental investment could be used as a remedial mechanism. Furthermore, children with higher developmental skills might enjoy more attending school, regardless of how beneficial it is for their human capital accumulation. Ignoring these potential correlations would lead to biased estimates.

To address the endogeneity of inputs, we take a control function approach. The main advantage of this method is that it allows for non-linear second stage equations and therefore, it can accommodate a CES production function. This method requires some assumptions about the timing and structure of the unobserved productivity components that we discuss in the remainder of this subsection.

Taking logarithm on both sides of equation 4 yields:

$$ln\ y^d_{igap} = \frac{1}{\rho^d} ln\big(\beta^d I^{d^\rho}_{ip} + (1-\beta^d) S^{d^\rho}_{ip}\big) + \gamma^d X_{ip} + \eta^d_{ip} \tag{5}$$

We assume that the unobserved productivity term η^d_{ip} has two independent components, one that is observed by parents throughout the experiment (π^d_{ip}) and one that is only observed after IDELA tests are administered (ν^d_{ip}). Formally:

$$\eta^d_{ip} = \pi^d_{ip} + \nu^d_{ip} \tag{6}$$

The choice of parental investment and school attendance that parents make is determined by the buget and time constraints of the household as well as parents preferences. Unfortunately, we cannot elicit these preferences and constrains from our data. Thus, instead of specifying a model for parental decisions, we follow Attanasio et al (Colombia) and note that all we need for identification is an exogenous variable that shifts each of our inputs. Therefore, we take a reduced form approach for the first stage.

In our first stage equations, the school attendance (S_{ip}) and parental investment (I_{ip}) for children i with parents p are defined as a function of the parent and child determinants of productivity (X_{ip}) as well as exogenous shifters $g(Z_{ip})$ and the productivity shocks that parents observe π_{ip}^d. Furthermore, we assume that G is additive separable in its three components and linear in X_{ip} and π_{ip}^d. We remain agnostic about how Z_{ip} affects both inputs. Formally:

$$I_{ip} = \kappa^I X_{ip} + g^I(Z_{ip}) + \sum_s \phi^{d,I}\pi_{ip}^d \tag{7}$$

$$S_{ip} = \kappa^S X_{ip} + g^S(Z_{ip}) + \sum_s \phi^{d,S}\pi_{ip}^d \tag{8}$$

Where $g(\cdot)$ is an unknown function of Z_{ip}. Because ν_{ip}^d is realized after the IDELA test is administered, it is uncorrelated with parental investment and school attendance. On the other hand, π_{ip}^d can potentially affect the input choices of parents and therefore we need to control for the potential bias it introduces in the second stage.

Define $\Pi^I = \sum_s \phi^{d,I}\pi_{ip}^d$ and $\Pi^S = \sum_s \phi^{d,}\pi_{ip}^d$ as the unobserved components of equations 13 and 14 respectively. Then, by including Π^I and Π^S in the second stage we control for the part of $\eta_{ip}^d = \pi_{ip}^d + \nu_{ip}^d$ that is correlated with the investment decision ($i.e.$ π_{ip}^d) and regain consistency. This yields:

$$ln\ y_{igap}^d = \frac{1}{\rho^d} ln\big(\beta^d I_{ip}^{d^\rho} + (1-\beta^d) S_{ip}^{d^\rho}\big) + \gamma^d X_{ip} + \phi^S \Pi^S + \phi^I \Pi^I + \xi_{ip} \tag{9}$$

Where ξ_{ip}^d is the residual of the projection of η_{ip}^d over Π^I and Π^S which by construction is uncorrelated with inputs.

8.3 Non-parametric instruments

To estimate the production function accounting for unobserved productivity we require instruments for both school attendance and parental investment. The eight treatment groups created by the intervention are a natural choice of instruments. Unfortunately, the correlation between treatment and inputs is not strong enough to rule out the finite sample concerns that may arise in the presence of weak instruments (Andrews et al. 2019). For this reason, we use a three stage procedure that relies on a non-parametric approximation to the expectation of both inputs conditional on treatment assignment.

In our sample the F-statistics for the hypothesis that the seven treatment types affected inputs range between 3.4 for parental investment and 6.2 for school attendance. While this is a statistically significant association, the correlation is not sufficiently strong to avoid the finite sample bias and spurious coverage rate of instrumental variable estimators when instruments are weak.

Amemiya (1974) showed that given an exogenous shifter Z of an endogenous variable X, the optimal instrument for X, in the sense of minimizing the mean squared error, is the conditional expectation $E[X|Z]$. In most settings, this is an unfeasible instrument since the specific relation between X and Z is unknown and thus, it cannot be estimated without imposing strong parametric or functional assumptions. To overcome this problem, Newey (1990) proposes procedures that relies on a non-parametric estimator for the conditional expectation in order to obtain an approximation to the optimal instrument that is later used in a second stage. We take a similar approach but taking advantage of categorical nature of the treatment indicators which allows us to obtain estimates for the conditional expectation using a fully interacted linear model.

In our setting, children cannot be accepted into preschool before they are at least three years old. Therefore, school attendance is mechanically correlated with age. Furthermore, it is natural that parental investment adjusts as children grow up and become less dependant on their parents. Additionally, we found that there are significant differences in attendance and parental investment between male and female children, these might be caused by the beliefs and expectations of parents which the intervention is likely to change. For the previous reasons, it is plausible that treatment effects differ across children by age and gender.

Age and gender could be correlated with unobserved productivity which implies we cannot directly use them to derive instruments. However, assignment to treatment is, by definition, uncorrelated with any observed or unobserved determinant of productivity. This means that after conditioning on age and gender, the interactions of both variables with treatment indicators are conditionally exogenous and therefore can be used as sources of variation for identification. This motivates our estimation procedure which relies on estimates for the conditional expectation of inputs obtained in a first stage and non-parametrically controlling for age and gender in the second and third stages.

To formalize the previous reasoning let Z be a vector comprised by the 32 indicators for all the possible treatment, age and gender combinations. Then, the conditional expectation of parental investment and school attendance given treatment, age and gender can be consistently estimated using linear regression of both inputs on Z:

$$I_{igat} = \omega_{I,gat} Z_{gat} + \zeta_i \tag{10}$$

$$S_{igat} = \omega_{S,gat} Z_{gat} + \zeta_i \tag{11}$$

In the previous equations, Z_{gat} is an indicator for gender g, age a and treatment t and ζ_i is the child level deviation from the conditional mean. Therefore, at the population level $\omega_{I,gat}$ and $\omega_{S,gat}$ will correspond to $E[I|age = a, gender = g, treatment = t]$ and $E[S|age = a, gender = g, treatment = t]$ respectively and the corresponding ordinary least squares (OLS) estimates will be consistent for these conditional expectations.

8.4 Estimation

Having estimates for the conditional expectation of inputs in hand, we proceed as in a standard control function approach. First, we estimate the residuals for both inputs which then we include as control variables in our final empirical equation. To account for the three step nature of our procedure, we compute boostraped confidence intervals and p-values for our estimates.

Let $Z_{ip} = (\hat{\omega}_{I,gat}, \hat{\omega}_{S,gat})$ be vector that comprises the OLS estimates obtained from equations 10 and 11. To use Z as a vector of instruments, we need $\omega_{I,gat}$ and $\omega_{S,gat}$ to be uncorrelated with the productivity shocks in equations 5 through 14. The two elements of Z are functions of age and gender, both of which might be correlated with unobserved productivity, therefore, we need to fully account for the effect of age and gender on both inputs and developmental outcomes.

In our study, the age of children ranges from 3 to 6 years old. Therefore, there are only 8 possible age-gender combinations. This means that it is feasible to fully saturate (*i.e.* include an indicator variable for each combination as a control) the production function as well as the reduced for equations for both inputs with respect for age and gender. This approach ensures that none of our estimates are contaminated by spurious correlations introduced by age or gender. Formally, we re-write equation 12 as:

$$ln\ y^{d}_{igap} = \frac{1}{\rho^{d}} ln\big(\beta^{d} I^{d^{\rho}}_{ip} + (1-\beta^{d}) S^{d^{\rho}}_{ip}\big) + \alpha_{ag} + \gamma^{d} X_{ip} + \phi^{S}\Pi^{S} + \phi^{I}\Pi^{I} + \xi_{ip} \tag{12}$$

Where y^{d}_{igap} is the developmental outcome of interest, the first term comprises the potentially non-linear effect of both parental investment I and school attendance S and the second term and third term comprise the total factors productivity, which is determined by an age-gender constant term α_{ag} plus other shifters X_{ip} such as baseline inputs, parental education and parental expectations. The fourth and fifth terms correspond to the control functions and finally ξ_{ip} is the remaining unobserved productivity. We estimate 12 using non-linear regression.

Similarly, we adjust equations 13 and 14 to accommodate $\hat{\omega}_{I,gat}$ and $\hat{\omega}_{S,gat}$ as instruments while fully controlling for age and gender, this yields:

$$I_{ip} = \alpha_{I,ag} + \kappa^{I} X_{ip} + \theta_{S,S}\hat{\omega}_{S,gat} + \theta_{S,I}\hat{\omega}_{I,gat} + \sum_{s} \phi^{d,I}\pi^{d}_{ip} \tag{13}$$

$$I_{ip} = \alpha_{S,ag} + \kappa^{S} X_{ip} + \theta_{I,S}\hat{\omega}_{S,gat} + \theta_{I,I}\hat{\omega}_{I,gat} + \sum_{s} \phi^{d,S}\pi^{d}_{ip} \tag{14}$$

Where $\alpha_{I,ag}$ and $\alpha_{S,ag}$ correspond to age-gender fixed effects and X_{ip} comprises other determinants of total factors productivity. As discussed above, $\hat{\omega}_{I,gat}$ and $\hat{\omega}_{S,gat}$ are estimates for the optimal instruments and the last term corresponds to the unobserved productivity that affects the choices of parental investment and school attendance. Equations 13 and 14 are estimated by ordinary least squares.

Our estimation procedure relies on three steps, one to approximate the conditional expectations, one to obtain the control function and a final stage in which the production function is estimated. This implies that if we analytically computed standard errors from the third stage, we would fail to account for the estimation error introduced in every step. For this reason, we compute bootstrap confidence intervals and p-values for our estimates. We conducted 200 bootstrap replications clustered at the community level.

8.5 Marginal product of school attendance and parental investment

Estimating a production function allows us to learn what is the effect on developmental outcomes of an increase in either parental investment or school attendance holding all else equal. We do this by calculating the marginal products implied by our specification of the production function.

Differences in inputs and observable characteristics imply that each child has a different marginal product from both inputs. To obtain an summary measure of productivity, we calculate the marginal products focusing on minority children at the mode of age and gender, the median of all other characteristics and inputs and assuming that unobservable productivity equals zero. Then, equations 4 and 12 imply that the marginal product of parental investment at median attendance $\bar{S}$ and investment $\bar{I}$ and the median of covariates $\bar{X}$ is:

$$MP(I;\bar{S},\bar{X}) = \Big(\beta^d \bar{I}^{\rho-1}\Big)\Big(\beta^d \bar{I}^{\rho} + (1-\beta^d)\bar{S}^{\rho}\Big)^{\frac{1-\rho}{\rho}} exp\Big(\alpha_{ag} + \gamma^d X\Big) \tag{15}$$

And for school attendnace the marginal product is:

$$MP(S;\bar{I},\bar{X}) = \Big(\beta^d \bar{S}^{\rho-1}\Big)\Big(\beta^d \bar{I}^{\rho} + (1-\beta^d)\bar{S}^{\rho}\Big)^{\frac{1-\rho}{\rho}} exp\Big(\alpha_{ag} + \gamma^d X\Big) \tag{16}$$

Point estimates for this quantities can be obtained using the delta method after estimation of the production function. Moreover, estimators obtained using the delta method are asymptotically normal which allows to obtain valid p-values and confidence regions using the bootstrap method.

8.6 Results

The results from estimating the production function of literacy, numeracy, motor and socioemotional skills are presented in tables 7 and 8. We find that the relative contribution of both inputs is approximately similar and that, for the children in our sample, parental investment and school attendance exhibit a high degree of substitability. This implies that both inputs play an important role in skill formation and thus, an intervention that promotes one at the expense of the other is unlikely to have a significant positive effect.

Table 7 presents our estimates for the contribution of both inputs to total production and the parameter that determines the rate of substitution ρ. Each column corresponds to a different developmental skill. The parameter ρ dictates the elasticity of substitution, we find that ρ is significantly higher than zero which implies that the production function exhibits a higher degree of substitability between inputs than a Cobb-Douglas production function. Note that values of ρ close to 1 imply almost perfect substitution. Regarding factor intensity, recall that the parameter β determines how intensive in parental investment the production function is; values of β below 0.5 imply that the production function is more intensive in school attendance than parental investment. We find that β ranges between 0.47 and 0.53 which suggests that both inputs are equally important. This is consistent with our experimental results where we found that the effect of the increase in school participation was offset by reductions in parental investment.

In addition to estimating the production function, we assess the contribution of each input. The second panel of table 7 shows our estimates for the marginal product of parental investment and school attendance for the median minority child in our sample calculated as described in the previous subsection. We find that the marginal product of both inputs is very similar in magnitude across all four skills, ranging between 0.25 standard deviations for literacy to 0.46 standard deviations in the case of socioemotional skills.

Table 8 presents how demographics and parental expectations relate to the total factor productivity. The first panel shows the linear coefficients for the different age-gender combinations in our sample. As it is reasonable to expect, we find that as age increases the cognitive and non-cognitive skills of children improve. Furthermore, we find that those children whose parents completed secondary and those who are expected to complete secondary themselves are significantly more productive. This confirms that parental background and beliefs are important determinants of child development that should be central when considering the goals and implications of child welfare policies.

One of the main goals of this study is to understand the differences between minority and majority children. All our specifications include an indicator for minority children, after controlling for demographics and background as well as inputs, we do not find significant remaining differences except for the case of literacy. This points to additional obstacles that minority children might face due to language and cultural barriers.

A plausible explanation for the observed differences across ethnicities is discrimination. Discrimination inside kindergartens is likely to undermine the effectiveness of school participation by creating additional obstacles for the school achievement of minority children. We explore this hypotheses by including our discrimination index and its interaction with the minority coefficient in our production function specification. We do find that children from communities in which there is discrimination against minority children tend to perform better but there is no differential effect for minority children within the same communities. This suggests that in discrimination at school might be as important for the academic performance of enrolled children (at least in the short term), however as table 3 shows, it reduces school participation which in turn affects developmental skills.

Finally, the last panel of table 8 shows the coefficients for our unobserved heterogeneity in productivity terms. Both are statistically significant which confirms that there is correlation between inputs and unobserved productivity. This panel also shows the first stage F-statistics which are 10.57 for parental investment and 10.05 for school attendance.

9 Contribution of inputs to differences in cognitive and non-cognitive skills

Estimating the production function of developmental skills allows us to better understand the contribution of parental investment and school participation to cognitive and non-cognitive skills formation. We ask to what extent the differences in inputs explain the gap in cognitive and non-cognitive skills. Finally, we assess the relative effectiveness of both inputs at reducing this gap and how this effectiveness is affected by the elasticity of substitution and initial input levels.

9.1 Decomposing the developmental skills gap

To asses what fraction of the gap in developmental skills is explained by inputs, we conduct an exercise in which we estimate what would be the observed skills if all other differences where eliminated. We find that differences in parental investment account for 33% (in the case of literacy) and up to 73% (in the case of socioemotional skills) of the observed gap while school attendance accounts for between 19% and 38% of the gap.

We start by calculating the predicted output levels at their observed inputs, demographics and background for every child in our sample. From here, we obtain the differences that our model predcits and can compare them to the ones we actually observe. Then, we repeat this exercise replacing the observed parental investment I_{ip} of child i with parents p by the average of the majority $\bar{I}^M$ and assuming that all children have the average productivity estimated among majority children $\bar{A}^M$. This allows us to obtain the difference in outcomes that our model attributes solely to the gap in school participation by comparing the mean of the obtain predictions for majority and minority children. In a similar manner, we replace the observed school attendance levels S_{ip} by the mean of the majority holding productivity fixed at $\bar{A}$ to obtain the difference in outcomes that our model attributes to parental investment. Finally, we hold both inputs constant at the majority means ($\bar{S}^M$ and $\bar{I}^M$) to obtain the difference that is attributable to differences in productivity.

Our results are shown in table 9, since differences in parental investment are more severe than in school participation, it explains a larger fraction of the observed differences. Furthermore, we find that 42% of the differences in literacy are explained by factors other than inputs. This is consistent with our findings from table 8 where we found that the indicator of being a minority children has a negative and significant coefficient. On the other hand, we find that difference sin productivity do not explain the observed differences in socioemotional and motor skills, where minority children might have a higher total factor productivity which partially compensates for the lower input levels.

9.2 Marginal products, complementarity and substitution

To conclude, we ask what would be the effect of interventions aimed at jointly increasing parental investment and school participation. We do this by calculating what would be the effect of a one standard deviation increase in school attendance at all possible levels of parental investment and viceversa; for every child in our sample and then computing the average for both majority and minority children. In this excercise, the lower input levels of minority children affect marginal productivity through two opposing mechanisms. The decreasing marginal product property of the C.E.S. production function implies that Roma and Turkish children have higher marginal productivity due to their lower initial levels, however, since inputs are not perfect substitutes, there are gains from complementarity which, given their lower input levels, are smaller for minority children.

We find that the effect of increasing school attendance is higher when it is accompanied by higher parental investment, this is depicted in Figure 5. A marginal increase in school attendance at the observed levels would lead to increases of 0.52, 0.89, 0.94 and 0.64 standard deviations in literacy, numeracy, socioemotional and motor skills respectively. However, if parental investment went from the minority average of 1.34 to the majority average of 1.64 the effects would instead be 0.57, 1.1, 1.2 and 0.79 respectively. In all cases we find that the higher levels of parental investment translate into higher marginal benefits for majority children than for Roma and Turkish.

In the case of increases in parental investment we find that the marginal gains (departing from the conditions observed in our data) are higher for minority than majority children. This is due to the fact that the initial levels of parental investment are considerably lower. While this is dampened by the higher school participation of minority children, it is not enough to reverse the effect. We find that a marginal increase in parental investment would lead to increases of 0.41, 0.73, 0.75 and 0.45 standard deviations in literacy, numeracy, socioemotional and motor skills respectively. However, if the attendance level went from the minority average of 1.32 to the majority average of 1.46 the marginal products would become 0.46, 0.84, 0.93 and 0.54 standard deviations respectively.

Table 7: Production function of developmental skills: Factors substitution, intensity parameters and marginal products.

Dependent variable	Literacy (1)	Numeracy (2)	Socioemotional (3)	Motor (4)
Substitution parameter				
ρ	0.43	0.44	0.38	0.41
	[0]	[0]	[0]	[0]
Factor intesity				
β	0.49	0.52	0.53	0.47
	[0]	[0]	[0]	[0]
Median marginal product				
School attendance	0.28	0.49	0.46	0.27
	[0]	[0]	[0]	[0]
Parental investment	0.25	0.4	0.35	0.27
	[0]	[0]	[0]	[0]
Control functions				
ν_{Sc} (Parental investment)	-0.47	-0.47	-0.49	-0.5
	[0]	[0]	[0]	[0]
First stage F-statistic	[10.57]	[10.57]	[10.57]	[10.57]
ν_{Pi} (School attendance)	-0.49	-0.53	-0.57	-0.47
	[0]	[0]	[0]	[0]
First stage F-statistic	[10.05]	[10.05]	[10.05]	[10.05]
Observations	4,863	4,863	4,863	4,863

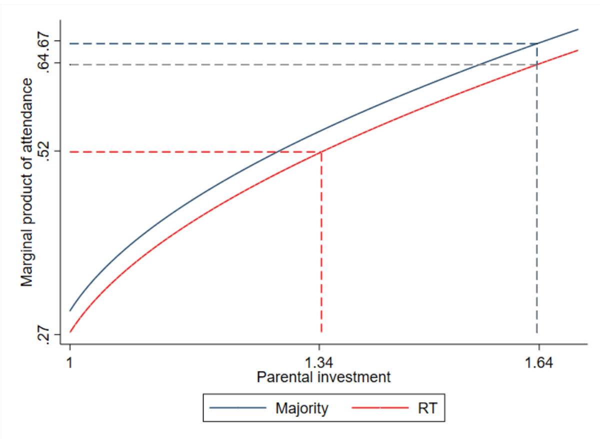
Table 8: Production function of developmental skills: Productivity shifters.

Dependent variable	Literacy (1)	Numeracy (2)	Socioemotional (3)	Motor (4)
Productivity parameters				
A (Constant)	-0.88	-0.21	-0.04	-0.96
	[0]	[0.04]	[0.7]	[0]
A_f (Female FE)	-0.08	-0.13	-0.08	0.09
	[0.41]	[0.11]	[0.25]	[0.27]
A_{A4} (Age=4 FE)	0.21	0.22	0.15	0.25
	[0]	[0]	[0.04]	[0]
A_{A5} (Age=5 FE)	0.6	0.54	0.4	0.65
	[0]	[0]	[0]	[0]
A_{A6} (Age=6 FE)	1	0.82	0.65	1.04
	[0]	[0]	[0]	[0]
$A_{f,A4}$ (Age=4#Female FE)	0.12	0.18	0.1	0.05
	[0.19]	[0.04]	[0.13]	[0.56]
$A_{f,A5}$ (Age=5#Female FE)	0.14	0.18	0.13	0.11
	[0.11]	[0.03]	[0.06]	[0.29]
$A_{f,A6}$ (Age=6#Female FE)	0.07	0.08	0.02	-0.03
	[0.44]	[0.35]	[0.81]	[0.72]
A_{RT} (Roma or Turkish FE)	-0.17	-0.09	-0.08	-0.01
	[0.05]	[0.12]	[0.26]	[0.92]
A_{PI} (Enrolled at Baseline)	-0.22	-0.26	-0.32	-0.25
	[0]	[0]	[0]	[0]
A_{PI} (Baseline Parent. Inv.)	-0.07	-0.07	-0.1	-0.1
	[0]	[0]	[0]	[0]
A_{PEd} (Parents completed sec.)	0.27	0.15	0.15	0.2
	[0]	[0]	[0]	[0]
A_{PEx} (Parents expect sec. ed.)	0.23	0.11	0.09	0.22
	[0]	[0]	[0]	[0]
A_{Disc} (Discrimination)	0.2	0.29	0.32	0.29
	[0.12]	[0.01]	[0.02]	[0.02]
$A_{Disc\#RT}$ (Discrimination # RT)	0.14	0.14	0.14	0.1
	[0.38]	[0.19]	[0.33]	[0.54]
Observations	4,863	4,863	4,863	4,863

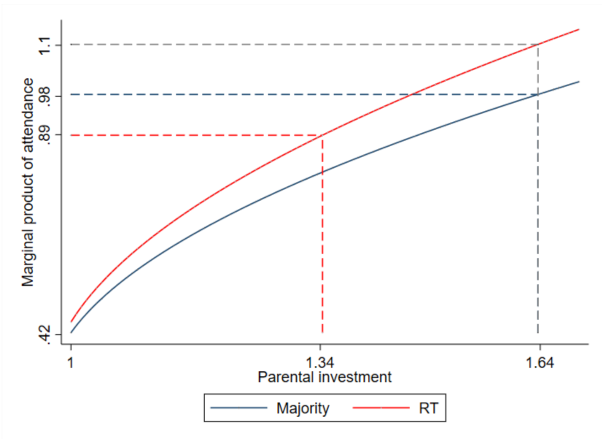
Table 9

	Observed difference	Predicted difference			
Developmental outcomes	Total	Total	School attendance	P. Investment	TFP
Literacy					
Mean	-0.402	-0.403	-0.077	-0.133	-0.168
Explained %		[100.25%]	[19.2%]	[33.1%]	[41.8%]
Numeracy					
Mean	-0.252	-0.253	-0.069	-0.146	-0.015
Explained %		[100.4%]	[27.4%]	[57.9%]	[6%]
Socioemotional					
Mean	-0.207	-0.207	-0.069	-0.153	0.037
Explained %		[100%]	[33.3%]	[73.9%]	[-17.9%]
Motor					
Explained %	-0.206	-0.206	-0.078	-0.138	0.032
		[100%]	[37.9%]	[67%]	[-15.5%]

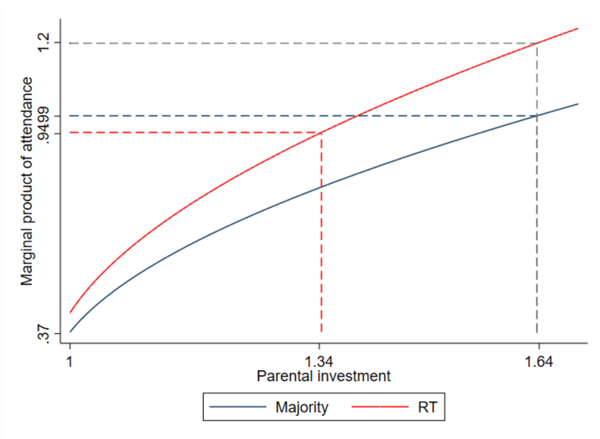
Figure 5: Marginal product of school attendance for Minority and Bulgarian children



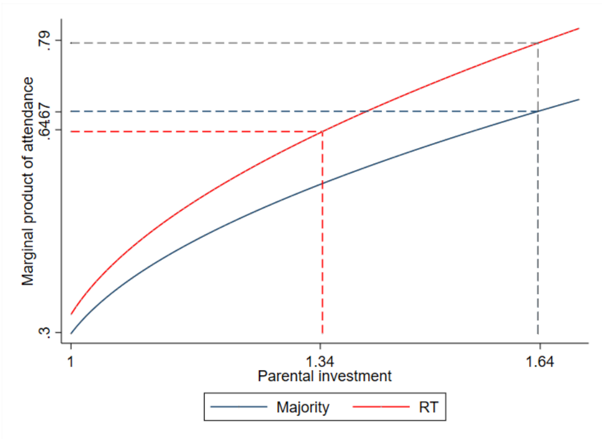
(a) Literacy



(b) Numeracy

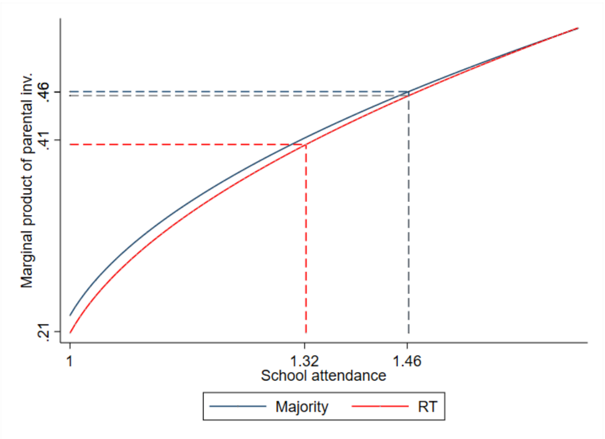


(c) Socioemotional

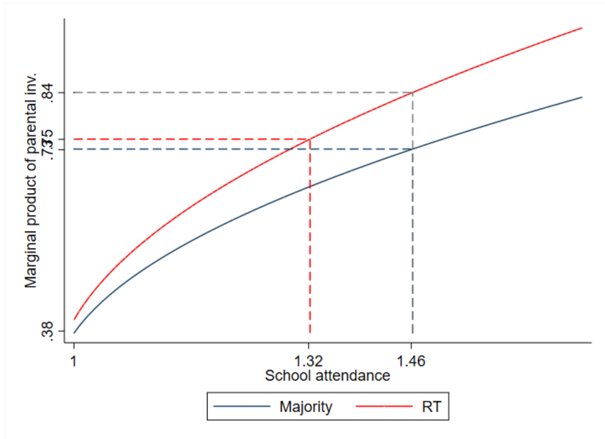


(d) Motor

Figure 6: Marginal product of parental investment for Minority and Bulgarian children



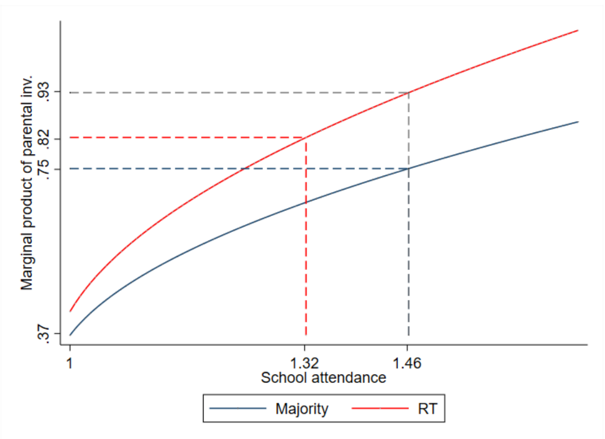
(a) Literacy



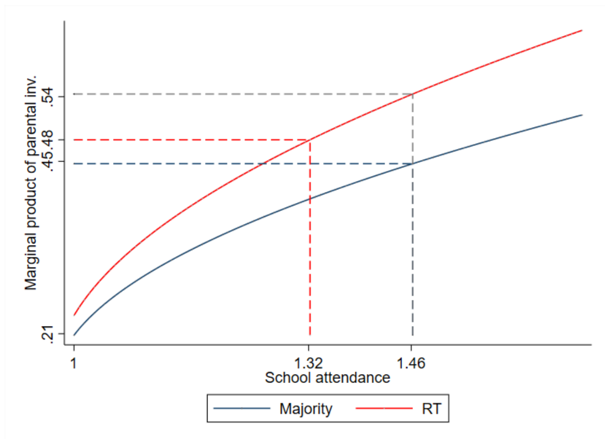
(b) Numeracy

(c) Socioemotional

(e) Motor



(d) Socioemotional



(f) Motor