Preschool or Parenting: Closing the Gap in Early Childhood Development Skills for the Roma Minority*

Paul Gertler
Elise Huillery
Joost de Laat
Ricardo Miranda

January 2025

Abstract

We investigate the contributions of preschool and parental stimulation activities at home to early child development skills in disadvantaged minority communities in Bulgaria. We present experimental evidence that combinations of information and monetary incentives increased preschool enrollment but also led to lower parental investments and ultimately lower child development. To rationalize these results, we estimate a child development skills production function using the experimental variation to identify the structural parameters. The estimated marginal products of both parental investments and preschool are positive and meaningful, with the contribution of parents being at least as large as the contribution of preschool, even for low-educated minority Roma parents. We also find that 2/3 of the child development gap between minority Roma children and majority Bularian children is explained by differences in parenting and 1/3 by differences in preschool. Our counterfactual calculations show that closing the gap in parental investment between minority and majority children would reduce the observed differences in literacy, numeracy, and socioemotional skills by 33, 58, 74 and 67%, respectively, while closing the gap in preschool attendance would reduce them by only 19, 27, 33 and 38%, respectively. The strong contribution of parents to child development favors policies that promote combinations of preschool participation and parental stimulation, even 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 Berkeley and NBER, gertler@berkeley.edu; Huillery: Université Paris Dauphine, elise.huillery@dauphine.psl.eu; de Laat: University of Utrecht, j.j.delaat@uu.nl; Miranda Montero: Duke University, 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. Funding study was provided by the World Bank's Srategic Impact Evaluation Fund (SIEF) (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. The project was approved by the IRB at the Paris School of Economics (Reference 2014-007).

1 Introduction

Large socioeconomic disparities in child development are evident when children enter primary school. Data from around the world show important gaps between socioeconomically advantaged and disadvantaged children at the early stages of life, both in developed (Pillas et al. 2014; Bradley et al. 2002; Kelly et al. 2011; Waldfogel et al. 2011) and in developing countries (Lu et al. 2020; López Boo 2016; Schady et al. 2015). These disparities emerge within the two first years of life (Fernald et al. 2012), widen between ages 3 and 5 (Kelly et al. 2011), and persist through adolescence and adulthood (Attanasio, Lopez-Boo, et al. 2024). Research on inequality in early childhood development also shows that the development penalty for disadvantaged children does not dissipate once they enter school (Schady et al. 2015), but rather expands over time, leading to a wide range of poor socioeconomic outcomes in adulthood (Grantham-McGregor et al. 2007). Parental stimulation at home and resources play an important role (Fernald et al. 2012; Akee et al. 2010), but these gradients persist even when controlling for nutrition and parenting (Naudeau et al. 2011).

To promote early child development and school readiness for the disadvantaged, preschool interventions are often considered promising programs for disadvantaged children. Programs such as Head Start, the Perry Preschool program, and the Carolina Abecedarian project, have shown high returns in the United States (Ludwig et al. 2007; J. J. Heckman et al. 2010; J. L. García et al. 2017; J. Heckman et al. 2013) and also in Denmark (Rossin-Slater et al. 2020), Argentina (Berlinski, Galiani, et al. 2009), Spain (Felfe et al. 2015), Germany (Hermes et al. 2024), and Ivory Coast (Yeo et al. 2024).

However, attending preschool does not always improve child development: a number of studies show that attending public preschool does not improve children'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), Cambodia (Bouguen et al. 2018), and Italy (Fort et al. 2020). These heterogeneous impacts of preschool on child outcomes may be related to variations in the quality of preschool across contexts and in the way disadvantaged children are treated by their teachers and peers, more or less inclusive or discriminatory.

In addition, heterogeneous impacts can also come from the substitution between preschool attendance and parenting at home. In fact, a large literature points to the crucial role

of parental stimulation which has consistently large positive impacts on child development (Schwartz et al. 2024, see Britto et al. 2017 and Attanasio, Cattan, and Meghir 2021 for literature reviews of the previous literature). Attending preschool should thus be seen as choosing one mode of education instead of another, so the outcome depends on how parental investment responds to preschool attendance as well as the relative productivity of preschool and parents in producing child development.

This paper investigates how parents and preschools shape child development in Bulgaria's concentrated Roma communities. Bulgaria has one of Europe's largest Roma populations, with the vast majority living in marginalized conditions. In 2014, 86% of Roma lived below the national at-risk-of-poverty threshold, compared to 22% of the general population (EU-MIDIS 2017). There was also a striking disparity in early education with only 66% of Roma children aged 4-6 attending preschool, versus 89% of ethnic Bulgarian children (EU-MIDIS 2017).

Given this gap, increasing Roma children's preschool participation presents a promising avenue for reducing subsequent educational and socio-economic disparities. To address this challenge, we conducted a large-scale randomized controlled trial to evaluate various interventions designed to boost preschool participation, including an information and mediation campaign, elimination of kindergarten fees, and additional financial incentives. The study design was an 8-arm cluster randomization across 236 poor communities with substantial Roma populations, encompassing 5,772 children aged 3-6 years. We measured the one-year impact of these interventions on three key outcomes: preschool participation, parental home stimulation, and child development.

The interventions significantly increased preschool participation among children not previously registered and substantially closed the gap in preschool enrollment between Roma and majority children. Across all treatment groups, enrollment rates exceeded the control group by 26 percentage points, while attendance rates were 0.40 standard deviations higher. Notably, we found similar effects across different combinations of interventions (information campaigns, free access, and additional financial incentives), suggesting that addressing any single barrier was sufficient to improve participation, with each approach showing comparable effectiveness.

However, our findings also revealed an unexpected trade-off: while the interventions increased preschool participation, they simultaneously led to a 0.25 standard deviation decrease in parental investment in educational activities with their children, suggesting that parents viewed preschool as a substitute for their own engagement with their children in educational activities. More concerning, children who were not registered in preschool at baseline showed negative developmental outcomes when exposed to the interventions. Across all treatment groups, these children scored 0.14 standard deviations lower in numeracy compared to the control group. Though not statistically significant, they also showed decreases of 0.11 standard deviations in both literacy and socio-emotional skills. These counterintuitive findings raise a critical question: why did increased preschool participation lead to reduced child development?

Two potential mechanisms warrant investigation: the reduction in parental investment and challenges that disadvantaged and minority children face in school. First, parental investment may be more important than expected, in particular with regard to highly disadvantaged parents such as Roma parents. Second, disadvantaged and minority children may get less out of preschool, i.e., have a lower marginal productivity of preschool because language and cultural differences may make it difficult to participate in school activities. Minority children may even suffer from discrimination by teachers and classmates. Lower parental investment may also reduce the marginal productivity of schooling as parents may be less likely to reinforce lessons taught in school at home.

We investigate the contributions of parents and preschool to child development by estimating a structural production function of child developmental skills (literacy, numeracy, and socio-emotional and motor) to quantify the degree of technical complementarity between school attendance and parental investments and their respective marginal products. Specifically, we specify a CES child development production function and use our experimental setting to construct instruments estimate its parameters. Specifically, we nonparametrically estimated the expected levels of school attendance and parental investment conditional on treatment assignment, age and gender, and then estimated the production function taking a control function approach using the estimated conditional expectations of the inputs as instruments.

The structural estimation results show that the marginal products of parental investment and of preschool attendance are both large and similar in magnitude for the median child. A one standard deviation increase in parental investment has approximately the same marginal product as a one standard deviation increase in school attendance. The results also show that increasing parental investment increases the marginal productivity of schooling. Thus, an increase in preschool attendance that crowds out parental investment is associated with a reduction in the marginal productivity of preschool. Hence, the observed reduction in children's skills was a result of both the direct effect of reduced parental investment and through its negative effect on the marginal productivity of schooling.

At observed levels of input utilization, the marginal product for minority parents is larger than for majority parents for numeracy, socio-emotional and motor skills, although not for literacy. This result is particularly important given that the median child has poor, low-educated, Roma parents. It means that even extremely disadvantaged parents can have large contributions to child development.

Overall, the 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 socio-emotional skills between by 33, 58, 74 and 67 percent respectively, whereas closing the gap in school attendance would reduce them by only 19, 27, 33 and 38 percent respectively. The implication of our findings is that the best policy to foster disadvantaged children' development would combine both access to formal education and parental investment in young children, which are rarely promoted together.

This paper makes several important contributions. First, we contribute to the literature on the demand for education. This literature already documents the separate effects of price and of information on the demand for education (on price Baker et al. 2008; Duflo et al. 2015; Duflo et al. 2021; Giacobino et al. 2024; S. García et al. 2022; Lucas et al. 2012, on information Jensen 2010). We advance this literature by analyzing how price and information interact across multiple price points, including both positive and negative values, to influence preschool demand. Our findings yield two novel insights. First, while confirming that eliminating direct preschool costs helps close the participation gap between disadvantaged and advantaged children, we provide the first empirical evidence that the information campaign

and free access act as substitutes rather than complements. This finding points to the importance of building proximity and togetherness between schools and minority parents who feel distant and discriminated in the education system Second, we demonstrate that negative prices through conditional cash transfers may not further stimulate demand once free access is provided and parents understand preschool benefits.

Our second major contribution enriches the literature on child development production functions. Recent advances in this field have illuminated complex interactions between children's health, cognitive skills, socio-emotional 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; Berlinski, Ferreyra, et al. 2024). We extend this literature by providing the first structural estimation of a production function that incorporates both preschool attendance and parental investment as inputs. Our analysis yields a crucial insight: parents and preschools are not perfect substitutes in producing child development outcomes. This finding has significant implications for education policy, suggesting that promoting preschool attendance without considering parental investment may yield suboptimal results. Our structural estimation of the relative productivity of parents and preschools complements the experimental work of Özler et al. 2018, who demonstrated that improved classroom quality alone does not enhance child outcomes without concurrent increases in parental inputs. These results underscore a clear policy directive: effective early childhood interventions must simultaneously promote both parental involvement and preschool participation.

Finally, our paper contributes to the vast literature on the importance of parents in early child development. While a strong consensus supports the effectiveness of parenting programs (Britto et al. 2017; Attanasio, Cattan, and Meghir 2021), our research reveals a crucial insight: disadvantaged parents substantially contribute to their children's human capital formation even without formal education or parenting support. This finding aligns with recent research showing that disadvantaged parents often underestimate the returns to educational investment, which adversely affects their investment decisions (Attanasio, Cunha, et al. 2019; Attanasio, Boneva, et al. 2019; Attanasio, Cattan, and Meghir 2021). Our results challenge conventional assumptions about parental capacity in disadvantaged communities: rather than

viewing disadvantaged parents primarily as barriers to their children's development, policy should recognize them as essential contributors to human capital formation and actively encourage their involvement.

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

2 The Institutional Context and Interventions

2.1 The Roma Minority and Education in Bulgaria

Roma are the largest ethnic minority in Europe, with a population of approximately 10-12 million. Their living conditions 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, and face unequal access to educational services. For example, 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. Although improving in some countries, the education gap starts early: the same survey reports that 53 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 Bulgaria, preschools serve 3 to 6-year-old children.[2] The vast majority of children enrolled in public early child education programs attend a full-day preschool, as opposed to half-day. 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. In 2011, the

[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

[2] Only the 5-6 year old groups are referred to as preschool, while the 3-4 year old groups are referred as kindergarten. For the sake of simplicity, this paper refers to all groups as preschool.

year with the most recent data on preschool enrollment for both Roma and non-Roma in Bulgaria before the start of this study, 45% of Roma children were estimated to be enrolled in preschool compared with 75% as a national average (The World Bank 2012).

Legal amendments from September 2010 made preschool education compulsory for all 5 and 6-year-old children, after which municipalities had two years to ensure that they were able to accommodate all children. For 5 and 6-year-old children, the educational services were thus made free of charge but kindergartens still charged fees to cover additional attendance costs such as food, school supplies, toilet paper, and required medical examinations. Furthermore, following a legal amendment in 2013, the monthly child allowance given to poor families[3] was made conditional on the child regularly attending preschool for children aged 5 and 6, unless this was impossible due to the condition of the child's health. On the contrary, 3-4 year old children were not concerned by the 2010 amendments and continued to pay fees for both educational services and attendance costs.

The cost of preschools is shared between the central government, local governments and families, from which the municipalities collect the 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. These, in turn, are channeled in full to preschools according to the number of children enrolled. The average central governmental financing of full-day kindergartens was BGN 1,702 per child in 2014 (across ages 3-6), with per child financing ranging from BGN 1,577 for each 3-4 year old in locations greater than 1,500 people, to BGN 1,822 for each 5-6 year old regardless of location. Local government funding covers infrastructure maintenance, as well as part of the running non-wage operating costs such as food, transportation, materials and similar items, through preschool fee revenues. The fees paid by families are determined by local governments and collected from primary caregivers directly by the preschools. Some municipalities set daily fees, others a monthly fee prorated for the actual days attended, others apply a combination of the two methods. In a few cases, fees differ monthly according to the cost of the food. According to official data, the average annual per child amount of preschool fees was estimated BGN 183 in 2014 (across ages 3-6).

[3]Families 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.

In September 2014, at the time this study began, whether these legal provisions had been successful in eliminating the costs and raising preschool rates among Roma children was not clear.

2.2 The Interventions

The primary goal of the impact evaluation was to assess ways by which to increase kindergarten enrollment and attendance, particularly among vulnerably minority children. The program Springboard for School Readiness (SSR) was jointly designed with the World Bank and the Trust for Social Achievement (thereafter, TSA) in Bulgaria as a demand side set of interventions to test three main reasons why children may not enroll and regularly attend preschool: first, parents may lack awareness around the benefits that preschool may provide around early stimulation of children and later success at school, 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 preschool attendance fees. Third, parents may be reluctant to enroll their children for other reasons such as 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 preschool 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 preschools were directly paid to the preschools 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 for whom the cost of preschool was more expensive.

Free access to full-day preschool + 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 preschool 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 preschool but fail at doing so because of time inconsistencies and lack of attention, a small amount may be sufficient to increase preschool participation. Again, the impact on parental behavior 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 preschool + 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 preschool participation than the small one in case parents have strong reasons why their kid does not attend preschool: they do not like preschool education, they have important opportunity costs of time going to preschool 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 behaviors for the 5-6 years old cannot be ruled out.

Information The intervention consisted of five public community meetings of 45 to 90 minutes each, organized at the local preschools, for a total of at least 6 hours over 10 months. The information campaign did not deal with parental skills and appropriate home-learning environment but provided parents with information about the benefits of preschool 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 preschool environment. In fact, perceived discrimination towards minority children by other families or teachers is likely to undermine school participation for Roma and Turkish children. Therefore, one session is a presentation by the teachers and another session is an "open house" were parents were invited to class. The sessions also promoted the interaction between parents and kindergarten officials. The concept was thus both a mediation and an 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 Table A1.

To implement the program, TSA partnered with 23 Bulgarian NGOs and collaborated with local preschools and municipal authorities. All interventions were implemented in full-day preschools. 16 of the 23 contracted NGOs were Roma-led. TSA trained each of the NGOs and each of full-day preschool principals participating in the financial interventions.

3 Methodology and Data

3.1 Experimental Design

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 authorities. 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 preschool 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 preschool. 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 preschool 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. All participant families provided informed consent before joining the study. Altogether 5,772 children were selected. Figure 1 summarizes the selection procedure.

A public lottery to assign the interventions was organized on June 30, 2014 in the presence of representatives of the Ministry of Education, the National Association of Municipalities in Bulgaria, mayors, directors of preschools, and NGOs. Using information from the baseline

survey, the random assignment was stratified by the implementing NGO and minimal distance in the number of available spots at preschool. 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 preschool, free preschool + BGN 7 bonus conditional on attendance, free preschool + 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 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.

3.2 Data

3.2.1 Data sources

A baseline survey was carried out in April-May 2014 to collect information on the communities, the preschools, and the households. In each community, the mayor, the preschool director, and a community leader were interviewed, as well as the 5,772 sampled households. The baseline survey captured information on demographics, education, literacy, numeracy, and employment of household members, as well as information on child rearing and 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 provided complete information from both parents and children for 4,863 of the 5,772 original sample of children, representing an attrition rate of 16%. In general, attrition was due to the fact that house-

holds and/or children were not available or not at home at the time the interviewer visited them. In addition to repeating baseline survey questions, the endline survey also included direct measures of child development using the International Development and Early Learning Assessment (hereafter, IDELA) developed by Save the Children. IDELA consists of 22 small exercises to assess four developmental areas: motor, literacy, numeracy, and socio-emotional development. 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. Save the Children provided training to the enumerators.[4] 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.

Finally, three unannounced attendance checks were carried out by OSI at the preschools during April-May 2015. During these checks, someone from the OSI team would make an unannounced visit to the preschool and verify which children were present and which children were not. Given the strategic nature of attendance in some treatment groups as well as the risk of social desirability, relying on an objective measure from unannounced visits was necessary to guarantee unbiased attendance estimates.

3.2.2 Variables of interest

Preschool participation We consider two measures of preschool 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. At baseline, enrollment is self-reported by parents, while at endline we also had access to administrative

[4]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, the tool has resulted in 2015 in a 22-item assessment that balances 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 IDELA's website https://idela-network.org/ for more information on the tool used for this project.

data.[5] Since enrollment is a minimal notion of school participation, we favor preschool attendance in our analyses, using 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. To facilitate comparison with measures of parental investment, we divided the attendance count by its standard deviation. This count of visits is the main measure of school participation throughout the study.

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, playing with toys, and naming, drawing and counting; 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 at baseline and endline.

Child development To measure developmental outcomes we used the IDELA tests administered at endline, designed to measure literacy, numeray, socioemotional and motor skills. This tool provides a simple way to measure the four skills of interest by calculating the score that each child obtained in the associated portion of the test. To facilitate comparison without ignoring the relative difficulty of each question, we divided all the scores by their sample standard deviation.

Socioeconomic variables Socio-demographic covariates were also collected on the household and household members, such as gender, age, income, an index of durable goods and amenities, perceived discrimination against Roma at preschool, 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

[5]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.

with treatment assignment. Parental expectations is a dummy indicating that parents expect their child to complete secondary education, because in our data most of the variation concentrates at this level of education. Similarly, for parental education, we just distinguish between parents who completed secondary education and those who did not. Finally, we construct an index for each household aimed at capturing discriminating practices against Roma and Turkish, which 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 other parents. 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 that we use in our analysis includes the replies from minority parents from other households that belong to the same community.

Table 1 presents a summary of our main variables and their measurement.

3.3 Baseline differences between minority and majority children

As our sample is composed by households living in communities with at least 10 households considered as Roma as per OSI's listing, majority households are likely not representative of the Bulgarian population. In this section, we present descriptive statistics on the sampled households and baseline differences between minority and majority children. Table 2 presents 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)

Panel A shows the means of the literacy, numeracy, socio-emotional and motor skills index at endline in the control group. We find significant differences in all four outcomes. The biggest difference concerns the literacy score: majority children scored on average 1,14 standard deviations, whereas minority children 0,85 standard deviations, which means a sizable gap of almost a third of a standard deviation. The majority-minority gaps in numeracy, socio-emotional skills, and motor skills amounted to respectively 23, 13, and 22 percent of a standard deviation. Majority children thus clearly outperformed minority children at endline

in the control group.

Panel B repeats the same exercise for baseline differences using all treatment and control groups. We find a significantly lower enrollment rate for minority children as well as consistently lower levels of parental involvement across all the surveyed activities—all differences being significant except for the playing and singing activities. The biggest difference in activities concerned reading books, which only 41% of minority parents reported having done in the last three days compared to 71% among majority parents. Therefore, both school participation and parental investment were highly differential according to ethnicity.

We find several plausible explanations for these differences. One is that minority parents have lower educational attainment: only 12% of minority parents completed secondary education compared to 48% of majority parents. Minority parents also report lower expectations regarding the educational accomplishment of their offspring, with 76% expecting their child to complete secondary compared to 90% of majority parents. We also find severe differences in income amounting to 41% of a standard deviation. Finally, 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. Our results illustrate the multiplicity of dimensions in which minority children face disadvantaged compared to majority ones, probably explaining the observed input differences which in turn contribute to observed developmental disadvantages.

These descriptive statistics also provide evidence that majority households in our sample are more disadvantaged than the general population in Bulgaria. In fact, only 48% of majority parents in our sample completed secondary education, compared to 74% in the total population aged 25 or more at the time of the survey in 2014.[6] Unemployment is also much more prevalent in our sample, with 30% of majority households compared to 11% in the general population in Bulgaria in 2014.[7] In our study, majority parents were thus also highly disadvantaged compared to the average household in Bulgaria, which should be kept in mind for the interpretation of our following analyses.

[6] https://data.worldbank.org/indicator/SE.SEC.CUAT.UP.ZS?locations=BG

[7] https://www.macrotrends.net/global-metrics/countries/BGR/bulgaria/unemployment-rate

We then investigate the extent to which socio-demographic and parental characteristics account for the ethnic differences in developmental inputs. We do this by estimating the following equation using a linear regression model:

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

In the previous equation, y_{ipcga} represents the variable of interest for child i of parents p 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 effects respectively. The matrix X_{ipc} entails parental characteristics and the community-level index that measures perceived discrimination toward Roma and Turkish children at school, as well as its interaction with a Roma and Turkish indicator. The term ϵ_i captures the remaining unobserved variation. As there is no guarantee that ϵ_i is uncorrelated with the regressors, our results merely shed light on correlations but should not be interpreted as causal relations.

Table 3 shows the results from estimating multiple versions of equation 1 by ordinary least squares. Column 1 shows that minority children are 17% less likely to be enrolled, and Column 2 that this difference is not attributable to demographics. Columns 3 and 4 show that both income and parental expectations are highly correlated with enrollment and capture about 7 percentage points of the observed difference while unemployment is not. Column 5 shows that, conditional on income and parental expectations, parental education is not significantly associated with enrollment. Column 6 shows that the observed differences in enrollment are highly correlated with perceived discrimination. Column 7 shows that children that receive higher parental investment are more likely to be enrolled. Columns 8 and 9 of table 3 show that minority children receive 0.31 less standard deviations of parental investment and that this difference is not linked to demographics. Columns 10 and 11 show that household income, labor participation, and parental expectations partially explain this gap, as well as parental education (Column 12). In column 13 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. Column 14 shows that, despite potentially spending less time at home, children who are enrolled receive more parental investment suggesting that parents who value

preschool are more likely to spend time with their children.

3.4 Estimation Strategy

The experimental design of the intervention allows us to estimate the average treatment effects by comparing outcomes in the control group and the different treatment groups. providing that attrition was random. Table 4 provides evidence that we cannot reject the null that the baseline characteristics of each treatment group are statistically identical to those of the control group.

To increase estimation precision, we estimate treatment effects adjusted for baseline covariates (age, gender, parental characteristics) using machine learning techniques to approximate an additively separable, potentially non-linear nuissance function following the approach developed by Chernozhukov et al. (2018) known as double/debiased machine learning. The outcomes of interest and treatment take-up are defined by the following two equations:

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

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

Where D_{it} indicates that child i was assigned to treatment t, 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 perceived 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. To estimate the partially linear model described by equations 2 and 3 we use the double/debiased machine learning algorithm of Chernozhukov et al. 2018 as implemented by the Stata 16 package *pdslasso* developed by Ahrens et al. 2020. 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.

Finally, since the intervention is more likely to affect children that were not enrolled,

we perform the analysis 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.

4 Experimental Results

4.1 Inputs and child development

Table 5 presents estimation results for the average treatment effects of the interventions on developmental inputs. Panel A 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 4 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 26 percentage points in enrollment (from a basis of 60% in the control group, hence a 43 percent increase), and a 0.40 standard deviation increase in attendance, among children that were not enrolled at baseline. On the contrary, parents reduced their investment as a response to the intervention: as columns 5 and 6 show, parental investment decrease by 0.25 standard deviations among children that were not enrolled at baseline.

Panel B of Table 5 reports the average effect of each specific treatment arm. We find consistent increases in enrollment and attendance across all treatment types. 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. However, when we test the equality of all treatment arms, p-values are always greater than 0.10 so we cannot reject the null that all treatments had similar effects on developmental inputs.

The reduction in the parental investment index is decomposed activity by activity in Table A2. Panel A shows that, on average, all activities were reduced (reading, telling stories, singing, playing, and naming). These reductions tend to be larger for children who were not enrolled at baseline. They are not independently statistically significant except for singing, which was reduced by 47 percent of a standard deviation for children who were not enrolled at baseline, as well as for playing, which was reduced by 21 percent of a standard

deviation for children who were already enrolled and 37 percent for children who were not enrolled at baseline. Panel B of Table A2 shows that many of the treatment-specific effects on the different activities were large and negative.

Finally, Table 6 reports estimates for the effect of the intervention on child development. In Panel A, we find a significant 0.14 standard deviation reduction in numeracy among children who were not enrolled at baseline, and a significant increase in motor skills among children who were already enrolled. For children who were not enrolled at baseline, for whom participation in preschool increased significantly, literacy skills did not improve—the coefficient is negative and non-significant, and the same applies to socio-emotional skills and motor skills. Desegregating the effects across treatment arms in Panel B reveals a similar pattern, with some marginally significant reductions in numeracy and literacy ranging between 0.14 and 0.17 standard deviations for children not enrolled at baseline, as well as some increases in motor skills among children already enrolled that range from 0.16 to 0.20 standard deviations.

It is therefore surprising that increased participation in preschool, which was the mere goal of the interventions, was associated with no improvement in child development—and even a reduction in numeracy skills. This counterintuitive finding challenges the common wisdom that disadvantaged children develop more rapidly in the formal education system than at home, and raises the critical question of explaining why increased preschool participation led to reduced child development. Two potential explanations need to be considered. One may be the observed decline in parental investment and the possibility that parental investment may be essential for child development, at least as much as preschool participation. Alternatively, disadvantaged children from poor communities may face structural challenges in the formal education system that may prevent them from benefiting from school participation, leading to negative marginal product of preschool attendance. In the rest of the paper, we investigate these explanations using a structural model to estimate the relative contributions of parents and preschool to child development. However, before that, we briefly explore the reasons why parents reduced their investment in response to the interventions.

4.2 The causes of the reduction in parental investment

Our experimental results shed some light on the causes of the reduction in parental investment. One possible explanation is that the intervention may have increased labor participation of the caregivers, reducing the time that they can dedicate to their children. This hypothesis is rejected. As Table A3 shows, the intervention had no effect on the income or income from labor, while labor participation marginally increased: the likelihood to be unemployed was reduced by only 1 percentage point from a basis of 40 percent among parents whose children were not enrolled at baseline. Such a small change in labor participation cannot explain the large reduction in parental investment. More generally, we find that parents who are employed invest *more* in activities with their child: Table 2 shows that majority households are less likely to be unemployed, and parental investment is significantly larger, compared to minority parents. Table 3 also shows that children whose household head is unemployed receive less parental investment on average. Given that participation to the labor market does not constitute a barrier to parental investment in child development and that the interventions did not change substantially labor participation, this is clearly not the reason why parents reduced their investment.

A second possible explanation is that preschool attendance may have reduced the time that children have to interact with their parents. In fact, given that the interventions encouraged participation to full-day preschool programs, one may worry that time left with parents was too small to permit parental engagement into developmental activities with their child. However, our data does not support this explanation either: Table 3 shows that children who attended preschool at baseline received more parental investment on average. Thus, from a descriptive point of view, the two inputs (parental investment and preschool participation) are positively correlated, which suggests that parents who choose preschool also engage into activities with their children more frequently. Table 2 illustrates this positive correlation by showing that majority children received more parental investment while at the same time participating more in preschool, compared to minority children. Therefore, we should not see the reduction in parental investment as a mechanical response to preschool attendance. On the contrary, the data shows that there is no incompatibility between parents doing activities with their child and the child attending preschool.

Finally, it could be that parents consider preschool as a substitute for parental investment and, if the intervention increased awareness about the benefits of preschool, they might have perceived that it was no longer necessary to spend as much time with their children to attain similar or better child development outcomes. We find some evidence in favor of this explanation. As Table A5 shows, the interventions changed some perceptions for parents whose child was not already enrolled at baseline. First, the interventions reduced perceived discrimination toward minority children at school, both by the other children and by teachers. Second, the interventions increased awareness about the benefits of preschool for secondary education completion (the coefficient is also positive for primary education, but somewhat smaller and thus not significant). These findings clearly indicate that parents in the treatment group perceived preschool as both more welcoming and more beneficial for their child than parents in the control group. The interventions thus led parents to perceive preschool as more productive in child development, which might have led them to believe that there was lower value of their own investment. Unfortunately we did not ask parents about their perception of their own role so we cannot directly test the hypothesis that they perceived their own contribution as relatively less beneficial than preschool. Presumably, the increase in perceived benefits of preschool led parents to perceive themselves as relatively less productive, which may explain why they reduced their own investment in educational activities in response to treatment.

5 The Child Development Production Function

To rationalize our experimental results, we estimate a production function of developmental skills that takes parental investment and school attendance as inputs. We estimate a separate production function for each of the four developmental outcomes, which we treat as outputs: literacy, numeracy, socioemotional, and motor skills. We find that, all else constant, increasing either input leads to improvements in development outcomes, but, for minority children, the marginal contribution of parental investment is larger than that of preschool attendance.

5.1 Model and identification

We assume that the production function of developmental skills belongs to the Constant Elasticity of Substitution (hereafter C.E.S.) class. We choose this parametrization of the production function because it allows us to be agnostic about the degree of complementarity between inputs but, unlike other parametrizations of production functions such as transcendental logarithmic (translog), the C.E.S. production function does not require higher order transformations of the input variables (which implies we do not need more instruments than inputs to address potential endogeneity) .

The effectiveness of preschool and parental investment might depend on observable parent and child characteristics such as parental educational attainment, parental aspirations, age, and gender, but also on unobservable characteristics such as child's and parents' ability. We assume that total factor productivity is a function of these observable and unobservable characteristics. Thus, the production function is described by the following equation:

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

Where $d \in \{Numeracy, Literacy, Motor, Socioemotional\}$ 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 an age-gender fixed effect α_{ag}, parent and child observable characteristics (X_{ip}) such as the education level and expectations of parents, baseline inputs, ethnicity and perceived discrimination, and an unobserved component η^d_{ip}.

Parental investment (I) and preschool attendance (S) enter the production function 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 investments while values close to zero represent a production function that is intensive in preschool 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 C.E.S. converges to the Cobb-Douglas production function.

Taking logarithm on both sides of equation 4 and rearranging yields:

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

It is plausible that unobserved productivity η^d_{ip} is correlated with preschool attendance and parental investment. For example, parents of more productive children who benefit more from preschool are more likely to enroll their children and, given the larger return, are also more likely to spend time with them. On the other hand, it could be that parental investment is used by parents as a remedial mechanism, and therefore, children with lower unobserved productivity receive more attention and investment from their parents. To account for the potential endogeneity of inputs with respect to unobserved productivity, we leverage the exogenous variation that arises from the randomized intervention.

Parental investment and preschool attendance choices are made before the IDELA test is administered. This means that parents and children choose inputs before y^d_{igap} is realized. Thus, we assume that η^d_{ip} is only partially observed by parents. Formally, we decompose η^d_{ip} as follows:

$$\eta^d_{ip} = \nu_i + \nu_p + \varepsilon^d_{ip} \tag{6}$$

Where ν_i and ν_p represent heterogeneity in children and parents' productivity that is common across developmental outcomes and observed by parents before the IDELA test is administered but not by the econometrician. ε_{ip} accounts for the remaining heterogeneity that is unobserved by both parents and researchers at the time that investment and attendance choices are made. We assume that inputs are independent from ε_{ip}, this is consistent with the assumption that parents cannot systematically predict productivity shocks that have not been realized at the time inputs are chosen.

Parents choose input levels based on their time and budget constraints, preferences, and productivity. Therefore, we model parental investment as functions of socio-demographic characteristics (which are likely to affect the constraints, productivity and preferences of children and their parents), unobserved productivity ν_i and ν_p, and the intervention which,

as we document in the previous section, affected the preferences of parents and children for preschool attendance and parental investment and the constrains they face. Formally:

$$I_{ip} = \gamma_{ag}^{I} + \kappa^{I} X_{ip} + g^{I}\left(D_{ip}\right) + \phi^{I}\nu_{i} + \psi^{I}\nu_{p} + \xi_{ip}^{I} \tag{7}$$

$$S_{ip} = \gamma_{ag}^{S} + \kappa^{S} X_{ip} + g^{S}\left(D_{ip}\right) + \phi^{S}\nu_{i} + \psi^{S}\nu_{p} + \xi_{ip}^{S} \tag{8}$$

For $L = I, S$ the term γ_{ag}^{L} represents a gender-age fixed effect, we also include the remaining determinants of productivity from equation, namely X_{ip} which comprises observed determinants plus ν_i and ν_p which account for the unobserved determinants that parents observe. ξ_{ip}^{L} accounts for remaining unobserved preferences for inputs which we assume to be uncorrelated with age, gender, X_{ip}, ν_i and ν_p.

To identify the model we proceed in two steps. First, we use equations 7 and 8 to recover $\phi^{I}\nu_{i} + \psi^{I}\nu_{p} + \xi_{ip}^{I}$ and $\phi^{S}\nu_{i} + \psi^{S}\nu_{p} + \xi_{ip}^{S}$. Then, we use these terms to construct a valid control function for the unobserved heterogeneity in equation 4. This control function allows us to recover the structural parameters β^{d}, ρ^{d}, and the productivity parameters α_{ag}^{d} and γ^{d}. We provide a detailed identification argument in Appendix B.

To asses the effectiveness of different hypothetical policies we are interested in the effect of increasing parental investment or preschool attendance under different counterfactual scenarios and how these differ between minority and majority children. From equation 4 we can obtain analytical expression for the marginal product of each input which we can evaluate at different input and productivity levels to approximate the distribution of marginal products in our sample.

5.2 Estimation

To identify the model, we need variation in parental investment and preschool attendance that is independent from observed and unobserved productivity. A natural choice would be the 8 different treatments from the randomized assignment. By construction, assignment to the 8 different treatments is uncorrelated with observable and unobservable child and parent characteristics. Furthermore, as we show in the previous section, the intervention affected both preschool attendance and parental investment. However, using treatment assignment

has two caveats: first, while the different treatments are correlated with both preschool attendance and parental investment, and this association is statistically significant, we find that the F statistic for the null hypothesis that the intervention had no effect is not larger than 10; this implies that finite sample estimates of the model might be biased **(CITE!!!)**. Second, finite sample bias due to weak correlation between inputs and the different treatment types might be exacerbated if we used indicators of all 8 treatment types as instruments due to the many instruments problem **(CITE!!!)**.

To address the previous concerns we construct two instruments derived from the interaction between assignment to the 8 different treatment types with age and gender indicators. Specifically, we estimate the expectation of both parental investment I and preschool attendance S conditional on age-gender-treatment cells, which we denote $\mu^I_{a,g,t}$ and $\mu^S_{a,g,t}$. Then, we use these conditional expectations as instruments for both inputs. As we showed in the previous section, age and gender are balanced both across treatment types and between majority and minority children; thus, they are plausibly uncorrelated with unobserved differences in productivity. Furthermore, within age-gender cells, assignment to treatment is random, therefore, conditional on age and gender, $\mu^I_{a,g,t}$ and $\mu^S_{a,g,t}$ are valid instruments for preschool attendance and parental investment. This approach follows Amemiya 1974 who proposes using the conditional expectation as an instrument and shows that, in the linear case which in our setting would correspond to a Cobb-Douglas production function, this instrument is optimal in the sense of minimizing mean-squared error and (CITE Jacknife IV papers).

Thus, we replace $g^I(D_{ip})$ and $g^S(D_{ip})$ in equations 7 and 8 with estimates for $\mu^I_{a,g,t}$ and $\mu^S_{a,g,t}$ which, since age, gender, and treatment assignment are all categorical variables, can be obtained using a saturated linear regression and interpreted as non-parametric estimates for the corresponding conditional expectations.

To account for potential estimation errors introduced by potential finite sample errors in the estimation of $\mu^I_{a,g,t}$ and $\mu^S_{a,g,t}$ or estimation of $\phi^I\nu_i + \psi^I\nu_p + \xi^I_{ip}$ and $\phi^S\nu_i + \psi^S\nu_p + \xi^S_{ip}$ we obtained Bootstrap confidence intervals and p-values for all our production function estimates. We conducted 250 Bootstrap replications.

5.3 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 the same; the parameter that dictates the relative contribution of parental investment (β^d) ranges between .47 for literacy and .5 in the case of socioemotional skills. Moreover, parental investment and preschool attendance exhibit a high degree of substitability; the substitution parameter ρ is larger than 0 which implies the production function exhibits less complementarity between inputs than a Cobb-Douglas production function, although we reject perfect substitution (i.e. $\rho = 1$). These results confirm that both inputs play a crucial role in early skill formation, therefore, an intervention that promotes one at the expense of the other is unlikely to have a significant positive effect. This is consistent with our experimental results, where we found that the benefits of increasing school participation were offset by reductions in parental investment.

The last panel of table 7 shows the coefficients for the unobserved heterogeneity terms that we recover from equations 7 and 8 . Both coefficients are statistically significant, which confirms that there is correlation between inputs and unobserved productivity.

Estimates for the parameters of equations 7 and 8 are shown in table 10. Recall that in these equations we include the expectation of the dependent variables (parental investment or preschool attendance) as regressors; therefore, the coefficients in this equation can be interpreted as the association between other observable household characteristics and deviations from the average among children in the same age-gender-treatment group. We find that, after conditioning on age, gender and treatment, minority children are more likely to be enrolled, although children who where enrolled at or belong to homes in which parents completed secondary education (both of which are more frequent among majority children) are also more likely to attend preschool. Similarly, we find that children with high baseline parental investment or whose parents expect them to complete secondary education receive more parental investment. Moreover, children from communities where there is discrimination toward minorities at school have attendance rates lower than majority children or minority children who attend schools with less discrimination. Finally, our two constructed instruments are highly correlated with their associated outcome (which, by construction, is expected), and

have coefficients that are not statistically different from one which supports interpreting estimates that result from this equation as deviations from the expected treatment response within age-gender cells.

Table 8 shows the productivity parameters associated to observable socio-demographic child and household characteristics. The first rows show the coefficients that correspond to the different age-gender combinations in our sample. Unsurprisingly, we find that as age increases the cognitive and non-cognitive skills of children improve. Furthermore, we find that for numeracy and socioemotional skills, female children are significantly more productive.

Our results show that parental background and expectations affect the productivity of children. Table 5 shows that children whose parents expect them to complete secondary education or whose parents completed it are more productive in the case of literacy, although, as column 3 shows, we find that children whose parents completed secondary exhibit lower productivity for socioemotional skills.

We find that children who were enrolled at baseline or had higher levels of parental investment at the start of the study are less productive than children who started with disadvantages. This highlights the importance of guaranteeing universal access to preschool and promoting parental investment among the entire population and the potential of such policies to alleviate the disadvantages that minority children face. To further understand the differences between minority and majority children, all our specifications include an indicator for minority children. We find that even after controlling for demographics and background characteristics as well as inputs, there are significant differences in productivity across all developmental outcomes. This finding suggests the existence of additional obstacles that minority children and parents face, which are not captured by our data. These obstacles may be rooted in broader systemic disadvantages that demographic, economic, and parental background factors are not sufficient to explain.

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 hypothesis 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 significant differential effect for minority children within the same communities. This suggests that 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.

5.4 Decomposing the developmental skills gap

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 in reducing this gap and how this effectiveness is affected by the elasticity of substitution at initial input levels.

We start by calculating the difference in output levels that we observed in the data in order to compare them to the ones our model predicts under different counterfactual scenarios. Consistent with our descriptive results, we find lower levels of literacy, numeracy, socioemotional and motor skills among minority children. To asses what fraction of these differences is explained by preschool attendance, we replace observed parental investment I_{ip} of child i with parents p with the majority average $\bar{I}^M$ and assume that all children have the same productivity, which we fix to the mean 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 larger than differences in school participation, it explains a larger fraction of the observed differences. We find that differences in parental investment account for 30% (in the case of literacy) and

up to 66% (in the case of socioemotional skills) of the observed gaps, while differences in school attendance account for between 17.5% and 33% of the gaps. Furthermore, we find that 50% of the differences in literacy are explained by factors other than inputs. On the other hand, we find that differences in 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.

5.5 Marginal products, complementarity and substitution

To conclude, we ask what the effect of interventions aimed at jointly increasing parental investment and school participation would be. 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 vice versa, the effect of a one-standard deviation increase in parental investment at all possible levels of school attendance, for every child in our sample, and then computing the average effects for both majority and minority children. Naturally, differences in productivity directly affect the marginal product of both inputs. Moreover, 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 input levels; however, this difference is dampened by the fact that inputs are not perfect substitutes, which implies that the gains from complementarity are smaller for minority children.

Figure 4 depicts the marginal product of school attendance as a function of parental investment levels for minority and majority children. The x-axis represents parental investment while the y-axis represents the associated marginal product. The blue line represents majority children, while the red line represents minority children. The red dotted line points to the observed average parental investment for minority children and the associated marginal product. The blue dotted line does the same, but for majority children. Finally, the gray line shows what would be the marginal product of preschool attendance for minority children if their level of parental investment coincided with that of their majority peers.

In the case of literacy, we find that the marginal product for majority children is 11% of a standard deviation larger, and that this difference would be reduced by 6% of a standard

deviation if parental investment levels were equalized. For numeracy we find that the marginal product of preschool for minority children is lower, but after equalizing parental investment, it would be 3% of a standard deviation larger. For motor and socioemotional skills we find that even at lower levels of parental investments, minority children benefit more from increases in preschool attendance; this is explained by diminishing returns and the fact that minority children are not less productive than their peers in the case of motor and socioemotional skills.

The curve that traces the marginal product of parental investment as a function of preschool attendance is depicted in figure 5. The patterns we find are qualitatively similar for both inputs. We find that minority children would benefit less from increases in parental investment in the case of literacy (3% of a standard deviation) and that these differences would be dampened but not eliminated if parental investment was equalized. For numeracy, our estimates reveal that minority children have a marginal product that is 2% of a standard deviation lower, but would be 3% larger than for majority children if preschool attendance among minority was as high as for majority children. For motor and socioemotional skills, the difference in marginal product ranges between 1% and 2% of a standard deviation and could increase up to 9% in the case of socioemotional skills if attendance levels were equalized to the majority average. This shows that promoting parental investment would increase the effectiveness of policies aimed at reducing the skills gap through preschool attendance.

6 Conclusion

The implementing partner, the Trust for Social Achievement (TSA), in collaboration with the research team, disseminated the initial impact evaluation findings widely across public stakeholders in Bulgaria through presentations, meetings, and television. In addition, TSA carried out a nationally representative survey to assess attitudes toward removal of kindergarten fees, launched an awareness campaign of the findings in 50 municipalities, prepared a series of reports on possible financing mechanisms for preschool, and organized a variety of meetings, roundtables, and conferences with stakeholders, including the Ombudsman and Members of Parliament.

In September 2020, Bulgaria's National Assembly (Parliament) passed legal amendments where, for the first time, state budget funds were set aside to support municipalities reduce or remove monthly attendance fees for full-day kindergartens Volen and Laat 2021. In April 2022, Parliament voted to allocate additional funds and made it a legal requirement preventing municipalities from charging any fees. In total, at least 244,000 children annually are estimated to benefit Volen 2023.

References

Ahrens, A., Hansen, C. B., and Schaffer, M. E. (2020). "lassopack: Model selection and prediction with regularized regression in Stata". In: *The Stata Journal* 20.1, pp. 176–235.

Akee, R. K., Copeland, W. E., Keeler, G., Angold, A., and Costello, E. J. (2010). "Parents' incomes and children's outcomes: a quasi-experiment using transfer payments from casino profits". In: *American Economic Journal: Applied Economics* 2.1, pp. 86–115.

Amemiya, T. (1974). "The nonlinear two-stage least-squares estimator". In: *Journal of econometrics* 2.2, pp. 105–110.

Attanasio, O., Bernal, R., Giannola, M., and Nores, M. (2020). *Child development in the early years: Parental investment and the changing dynamics of different dimensions*. Tech. rep. National Bureau of Economic Research.

Attanasio, O., Boneva, T., and Rauh, C. (2019). *Parental beliefs about returns to different types of investments in school children*. Tech. rep. National Bureau of Economic Research.

Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, C., and Rubio-Codina, M. (2020). "Estimating the production function for human capital: results from a randomized controlled trial in Colombia". In: *American Economic Review* 110.1, pp. 48–85.

Attanasio, O., Cattan, S., and Meghir, C. (2021). "Early Childhood Development, Human Capital, and Poverty". In: *Annual Review of Economics* 14.

Attanasio, O., Cunha, F., and Jervis, P. (2019). *Subjective parental beliefs. their measurement and role*. Tech. rep. National Bureau of Economic Research.

Attanasio, O., Lopez-Boo, F., Perez-Lopez, D., and Reynolds, S. A. (2024). "Inequality in the early years in LAC: A comparative study of size, persistence, and policies". In: *LACIR Series* 132.

Attanasio, O., Meghir, C., and Nix, E. (2020). "Human capital development and parental investment in India". In: *The Review of Economic Studies* 87.6, pp. 2511–2541.

Baker, M., Gruber, J., and Milligan, K. (2008). "Universal child care, maternal labor supply, and family well-being". In: *Journal of political Economy* 116.4, pp. 709–745.

Berlinski, S., Ferreyra, M. M., Flabbi, L., and Martin, J. D. (2024). "Child care markets, parental labor supply, and child development". In: *Journal of Political Economy* Forthcoming.

Berlinski, S., Galiani, S., and Gertler, P. (2009). "The effect of pre-primary education on primary school performance". In: *Journal of public Economics* 93.1-2, pp. 219–234.

Bouguen, A., Filmer, D., Macours, K., and Naudeau, S. (2018). "Preschool and parental response in a second best world evidence from a school construction experiment". In: *Journal of human Resources* 53.2, pp. 474–512.

Bradley, R. H. and Corwyn, R. F. (2002). "Socioeconomic status and child development". In: *Annual review of psychology* 53.1, pp. 371–399.

Britto, P. R., Lye, S. J., Proulx, K., Yousafzai, A. K., Matthews, S. G., Vaivada, T., Perez-Escamilla, R., Rao, N., Ip, P., Fernald, L. C., et al. (2017). "Nurturing care: promoting early childhood development". In: *The Lancet* 389.10064, pp. 91–102.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). *Double/debiased machine learning for treatment and structural parameters*.

Clarke, D., Romano, J. P., and Wolf, M. (2020). "The Romano–Wolf multiple-hypothesis correction in Stata". In: *The Stata Journal* 20.4, pp. 812–843.

Cunha, F., Heckman, J. J., and Schennach, S. M. (2010). "Estimating the technology of cognitive and noncognitive skill formation". In: *Econometrica* 78.3, pp. 883–931.

Del Boca, D., Flinn, C., and Wiswall, M. (2014). "Household choices and child development". In: *Review of Economic Studies* 81.1, pp. 137–185.

Drange, N., Havnes, T., and Sandsør, A. M. (2016). "Kindergarten for all: Long run effects of a universal intervention". In: *Economics of Education Review* 53, pp. 164–181.

Duflo, E., Dupas, P., and Kremer, M. (2015). "Education, HIV, and early fertility: Experimental evidence from Kenya". In: *American Economic Review* 105.9, pp. 2757–97.

— (2021). *The impact of free secondary education: Experimental evidence from Ghana*. Tech. rep. National Bureau of Economic Research.

Felfe, C., Nollenberger, N., and Rodríguez-Planas, N. (2015). "Can't buy mommy's love? Universal childcare and children's long-term cognitive development". In: *Journal of population economics* 28.2, pp. 393–422.

Fernald, L. C., Kariger, P., Hidrobo, M., and Gertler, P. J. (2012). "Socioeconomic gradients in child development in very young children: Evidence from India, Indonesia, Peru, and Senegal". In: *Proceedings of the National Academy of Sciences* 109.supplement_2, pp. 17273–17280.

Fort, M., Ichino, A., and Zanella, G. (2020). "Cognitive and noncognitive costs of day care at age 0–2 for children in advantaged families". In: *Journal of Political Economy* 128.1, pp. 158–205.

García, J. L., Heckman, J. J., Leaf, D. E., and Prados, M. J. (2017). *The life-cycle benefits of an influential early childhood program*. Tech. rep. National Bureau of Economic Research.

García, S. and Saavedra, J. (2022). "Conditional Cash Transfers for Education". In.

Giacobino, H., Huillery, E., Michel, B., and Sage, M. (2024). "Schoolgirls, not brides: Education as a shield against child marriage". In: *American Economic Journal: Applied Economics* 16.4, pp. 109–143.

Goux, D. and Maurin, E. (2010). "Public school availability for two-year olds and mothers' labour supply". In: *Labour Economics* 17.6, pp. 951–962.

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., Strupp, B., Group, I. C. D. S., et al. (2007). "Developmental potential in the first 5 years for children in developing countries". In: *The lancet* 369.9555, pp. 60–70.

Gupta, N. D. and Simonsen, M. (2010). "Non-cognitive child outcomes and universal high quality child care". In: *Journal of Public Economics* 94.1-2, pp. 30–43.

Heckman, J., Pinto, R., and Savelyev, P. (2013). "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes". In: *American Economic Review* 103.6, pp. 2052–86.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., and Yavitz, A. (2010). "The rate of return to the HighScope Perry Preschool Program". In: *Journal of public Economics* 94.1-2, pp. 114–128.

Hermes, H., Lergetporer, P., Peter, F., and Wiederhold, S. (2024). "Application barriers and the socioeconomic gap in child care enrollment". In: *Journal of the European Economic Association* forthcoming.

Jensen, R. (2010). "The (perceived) returns to education and the demand for schooling". In: *The Quarterly Journal of Economics* 125.2, pp. 515–548.

Kelly, Y., Sacker, A., Del Bono, E., Francesconi, M., and Marmot, M. (2011). "What role for the home learning environment and parenting in reducing the socioeconomic gradient in child development? Findings from the Millennium Cohort Study". In: *Archives of disease in childhood* 96.9, pp. 832–837.

López Boo, F. (2016). "Socio-economic status and early childhood cognitive skills: A mediation analysis using the Young Lives panel". In: *International Journal of Behavioral Development* 40.6, pp. 500–508.

Lu, C., Cuartas, J., Fink, G., McCoy, D., Liu, K., Li, Z., Daelmans, B., and Richter, L. (2020). "Inequalities in early childhood care and development in low/middle-income countries: 2010–2018". In: *BMJ Global Health* 5.2, e002314.

Lucas, A. M. and Mbiti, I. M. (2012). "Access, sorting, and achievement: The short-run effects of free primary education in Kenya". In: *American Economic Journal: Applied Economics* 4.4, pp. 226–53.

Ludwig, J. and Miller, D. L. (2007). "Does Head Start improve children's life chances? Evidence from a regression discontinuity design". In: *The Quarterly journal of economics* 122.1, pp. 159–208.

EU-MIDIS, I. (2017). "Second European Union Minorities and Discrimination Survey Main results". In: *Luxembourg, Publications Office of the European Union*.

Naudeau, S., Martinez, S., Premand, P., and Filmer, D. (2011). "Cognitive development among young children in low-income countries". In: *No small matter: The impact of poverty, shocks, and human capital investments in early childhood development*, pp. 9–50.

Özler, B., Fernald, L. C., Kariger, P., McConnell, C., Neuman, M., and Fraga, E. (2018). "Combining pre-school teacher training with parenting education: A cluster-randomized controlled trial". In: *Journal of Development Economics* 133, pp. 448–467.

Pillas, D., Marmot, M., Naicker, K., Goldblatt, P., Morrison, J., and Pikhart, H. (2014). "Social inequalities in early childhood health and development: a European-wide systematic review". In: *Pediatric research* 76.5, pp. 418–424.

Romano, J. P. and Wolf, M. (2005). "Exact and approximate stepdown methods for multiple hypothesis testing". In: *Journal of the American Statistical Association* 100.469, pp. 94–108.

Rossin-Slater, M. and Wust, M. (2020). "What is the added value of preschool for poor children? Long-term and intergenerational impacts and interactions with an infant health intervention". In: *American Economic Journal: Applied Economics* 12.3, pp. 255–86.

Schady, N., Behrman, J., Araujo, M. C., Azuero, R., Bernal, R., Bravo, D., Lopez-Boo, F., Macours, K., Marshall, D., Paxson, C., et al. (2015). "Wealth gradients in early childhood cognitive development in five Latin American countries". In: *Journal of Human Resources* 50.2, pp. 446–463.

Schwartz, K., Michael, D., Torossian, L., Hajal, D., Yoshikawa, H., Abdulrazzak, S., Youssef, J., Sloane, P., Hashwe, S., Foulds, K., et al. (2024). "Leveraging Caregivers to Provide Remote Early Childhood Education in Hard-to-Access Settings in Lebanon: Impacts From a Randomized Controlled Trial". In: *Journal of Research on Educational Effectiveness*, pp. 1–31.

The World Bank (2012). *Toward an equal start: Closing the early learning gap for Roma children in Eastern Europe*. World Bank.

Volen, E. (2023). "Now that kindergarten is free of charge: laying the foundations for future pre-school policy change in Bulgaria". In: *Frontiers in Education*, pp. 1–5.

Volen, E. and Laat, J. de (2021). "Building Evidence for Pre-school Policy Change in Bulgaria". In: *Frontiers in Public Health* 9, pp. 1–8.

Waldfogel, J. and Washbrook, E. (2011). "Early years policy". In: *Child development research* 2011.

Yeo, A., Maiga, S., and Michel, B. (2024). *Should LMICs Build Preschools or Improve Primary Education? Experimental Evidence from Côte d'Ivoire*. Tech. rep. HAL.

Figures

Figure 1: Sample selection procedure

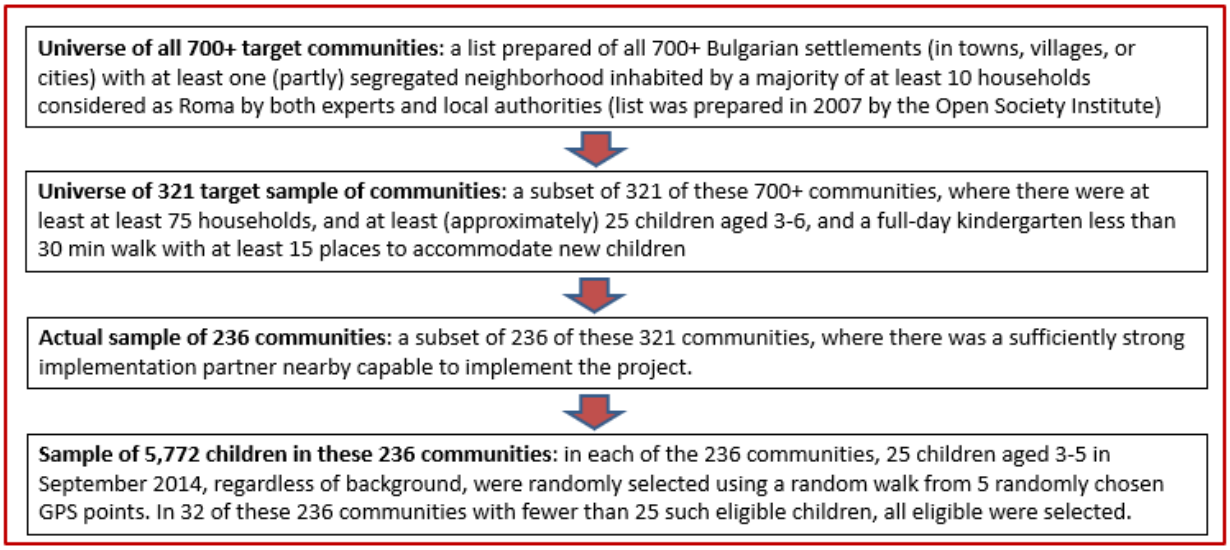


Figure 2: Distribution of treatment arms across selected communities

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

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

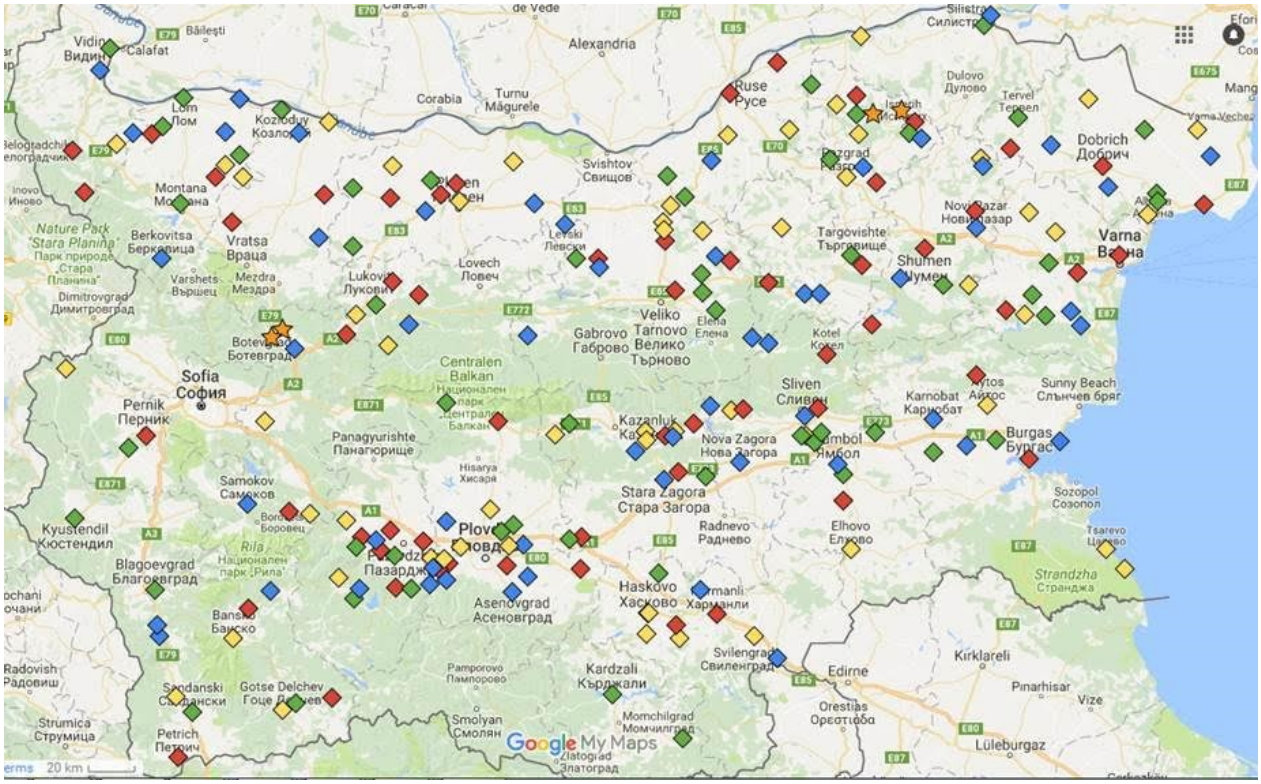
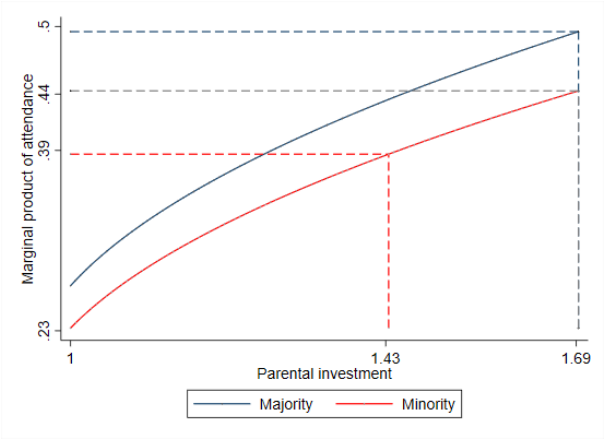
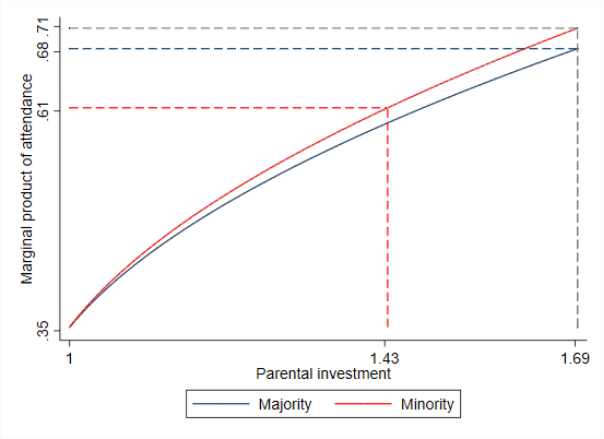


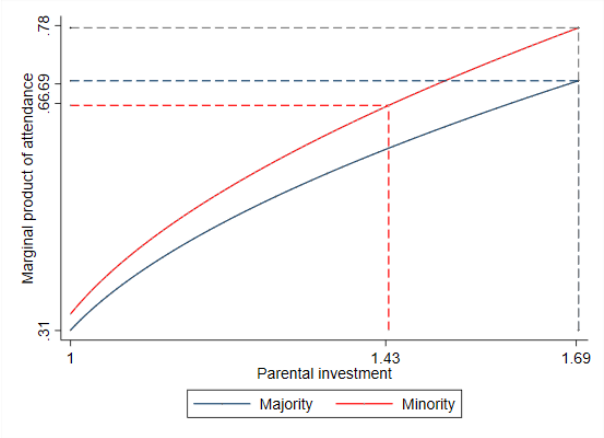
Figure 4: Marginal product of school attendance for Minority and Bulgarian children



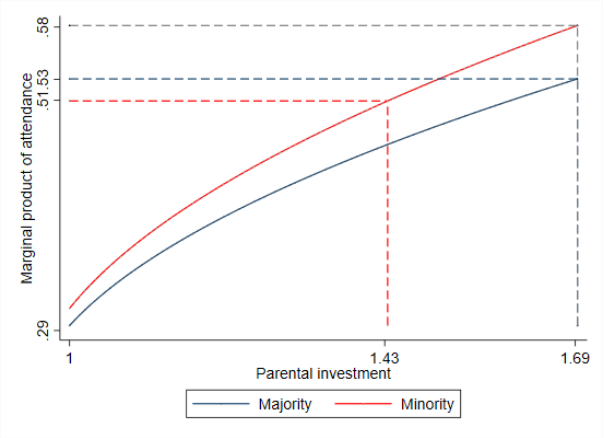
(a) Literacy



(b) Numeracy

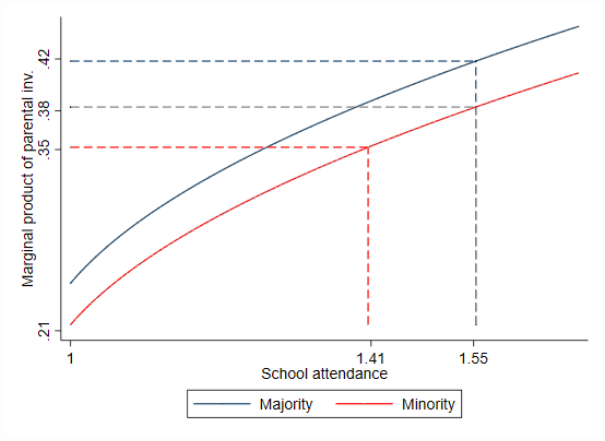


(c) Socioemotional

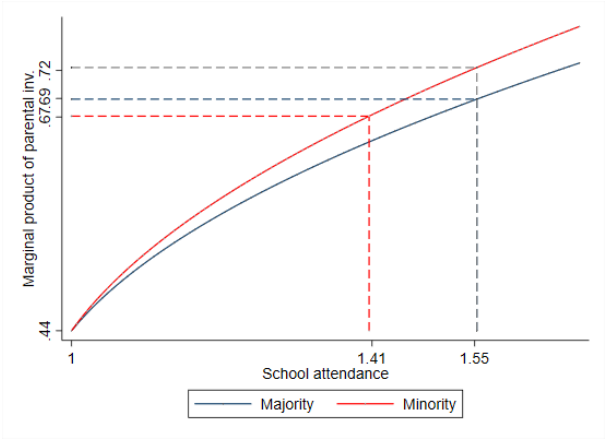


(d) Motor

Figure 5: Marginal product of parental investment for Minority and Bulgarian children



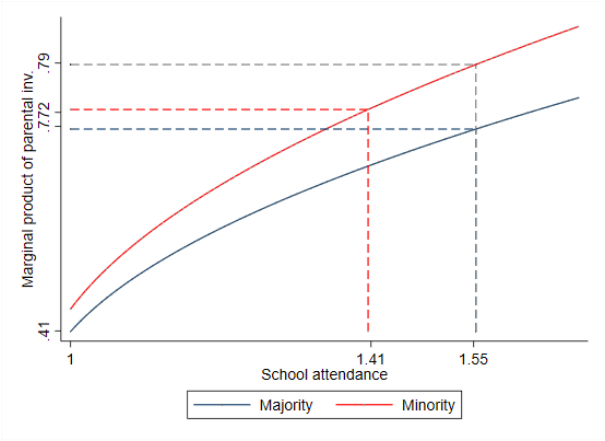
(a) Literacy



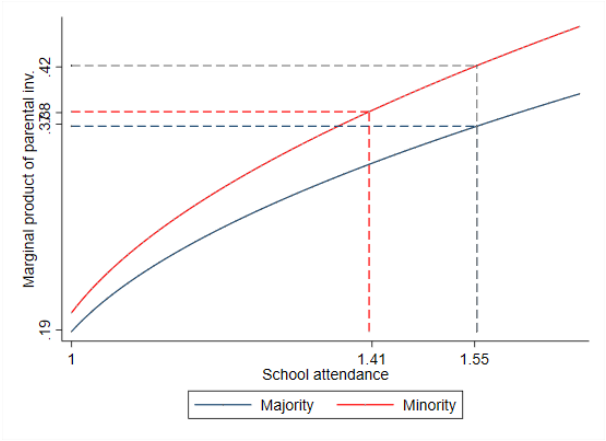
(b) Numeracy

(c) Socioemotional

(e) Motor



(d) Socioemotional



(f) Motor

Tables

Table 1: Variable measurement and construction

Variable	Data source	Type	Construction
Developmental skills			
Literacy	IDELA test	Z-score	Standardized sum of the answers to the IDELA questions aimed at measuring literacy
Numeracy	IDELA test	Z-score	Standardized sum of the answers to the IDELA questions aimed at measuring numeracy
Socioemotional	IDELA test	Z-score	Standardized sum of the answers to the IDELA questions aimed at measuring socioemotional skills
Motor	IDELA test	Z-score	Standardized sum 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	Standardized sum of indicators of the parents spending time reading, playing, naming, singing, telling stories and drawing with their kids
Endline investment	Baseline survey	Z-score	Standardized sum 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			
Household perception	Endline survey	Z-scores	Standardized sum of indicators of the parents percieving discrimination toward minority children from teachers and other parents
Minority community perception	Endline survey	Community avg. of Z-scores	Community average of minority households' percieved discrimination index (Majority households excluded)
Community perception	Endline survey	Community avg. of Z-scores	Community average of households' percieved discrimination index (Majority households included

Table 2: Differences between majority and minority households

	Majority	Roma or Turkish	H0: Majority=Roma or Turkish F-statistic	p-value	RW p-value
Panel A: Control group differences in developmental outcomes					
Literacy	1.14	0.85	13.53	0	0.01
Numeracy	1.67	1.44	9.82	0	0.01
Socioemocional skills	1.47	1.33	4.85	0.04	0.01
Motor skills	1.12	0.9	6.82	0.01	0.01
Observations per ethnicity	148	483			
Observations		631			

	Majority	Roma or Turkish	H0: Majority=Roma or Turkish F-statistic	p-value	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	0.01
Parental investment					
Read books	0.71	0.41	119.63	0	0.01
Told story	0.79	0.62	38.9	0	0.01
Sang	0.78	0.71	6.04	0.01	0.11
Went out	0.95	0.91	10.45	0	0.01
Played	0.79	0.74	3.56	0.06	0.16
Draw	0.78	0.64	24.67	0	0.01
Summary Index	1.68	1.37	37.39	0	0.01
Demographics					
Female	0.48	0.48	0.06	0.81	0.92
Age	4.76	4.75	0.09	0.77	0.92
Household characteristics					
H. Head Unemployed	0.3	0.42	18.41	0	0.01
H. Size	4.49	4.69	6.64	0.01	0.1
Income	6.02	5.61	59.49	0	0.01
Parental education					
Completed Kindergarden	0.2	0.37	183.04	0	0.01
Completed Secondary	0.48	0.12	157.78	0	0.01
Parental aspirations					
Age to stop education	21.73	19.31	116.53	0	0.01
Expects child to complete secondary	0.9	0.76	61.85	0	0.01
Percieved discrimination					
Household percievs disc. Toward RT at school (Ind.)	0.19	0.23	5.85	0.02	0.12
RT Community percievs disc. Toward RT at school (Ind.)	0.77	0.89	4.52	0.03	0.12
Community percievs disc. Toward RT at school (Ind.)	1.37	1.05	62.36	0	0.01
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)	(13)	(14)
Dependent variable	Baseline enrollment							Baseline parental investment						
Roma or Turkish	-0.17***	-0.17***	-0.15***	-0.10***	-0.09***	0.10***	0.10***	-0.31***	-0.31***	-0.26***	-0.18***	-0.16***	-0.01	-0.01
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.10)	(0.10)
log income			0.05***	0.04***	0.03***	0.03***	0.03***			0.06***	0.05***	0.04**	0.04**	0.04**
			(0.01)	(0.01)	(0.01)	(0.01)	(0.01)			(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
HH unemployed			0.00	0.01	0.01	0.01	0.01			-0.17***	-0.15***	-0.15***	-0.14***	-0.14***
			(0.02)	(0.02)	(0.02)	(0.01)	(0.01)			(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Parents expect secondary completion				0.17***	0.16***	0.11***	0.11***				0.26***	0.23***	0.24***	0.23***
				(0.02)	(0.02)	(0.02)	(0.02)				(0.04)	(0.04)	(0.04)	(0.04)
Household h. completed secondary					0.04	0.03	0.02					0.24***	0.24***	0.24***
					(0.03)	(0.02)	(0.02)					(0.05)	(0.05)	(0.05)
RT com. perceives disc. at school						-0.31***	-0.31***						0.13	0.15
						(0.04)	(0.04)						(0.09)	(0.09)
RT com. perceives disc. at school#RT						-0.20***	-0.19***						-0.19*	-0.17*
						(0.04)	(0.04)						(0.10)	(0.10)
Baseline parental investment							0.02**							
							(0.01)							
Baseline Enrollment														0.06**
														(0.03)
Constant	0.78***	0.53***	0.24***	0.22**	0.19**	0.48***	0.47***	1.68***	1.66***	1.31***	1.27***	1.10***	1.00***	0.97***
	(0.02)	(0.04)	(0.09)	(0.09)	(0.09)	(0.06)	(0.06)	(0.05)	(0.07)	(0.14)	(0.14)	(0.14)	(0.17)	(0.17)
Age and Gender FE	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Mean of the dependent variable	0.649	0.649	0.649	0.649	0.649	0.649	0.649	1.442	1.442	1.442	1.442	1.442	1.442	1.442
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	4,863	4,863
R-squared	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.06	0.06	0.06	0.06	0.06	0.06	0.06

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.

Table 4: Table of balance p-values

	Information	Free access	Free access + Info.	7 lev	7 lev + Info.	20 lev	20 lev + Info.
Developmental outcomes							
Literacy	1	1	1	1	1	1	1
Numeracy	0.74	0.35	0.49	1.00	0.99	0.66	0.54
Socioemotional sills	0.99	1.00	0.97	1.00	1.00	1.00	0.99
Motor skills	0.99	0.99	1.00	1.00	0.98	1.00	0.74
Kindergarden enrollment							
Registered at baseline	0.98	1.00	1.00	1.00	1.00	1.00	1.00
Parental investment							
Read books	0.87	1.00	0.94	1.00	1.00	0.85	1.00
Told story	0.89	0.98	1.00	1.00	1.00	0.79	0.93
Sang	0.96	1.00	1.00	1.00	1.00	1.00	1.00
Went out	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Played	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Draw	1.00	1.00	0.97	0.93	1.00	1.00	1.00
Summary index	0.96	1.00	1.00	1.00	1.00	1.00	1.00
Demographics							
Female	0.90	1.00	1.00	1.00	1.00	0.98	0.96
Age	1.00	1.00	0.94	1.00	0.99	0.98	1.00
Household characteristics							
H.Head unemployed	0.80	1.00	0.98	1.00	1.00	1.00	1.00
H. Size	0.55	0.99	0.89	0.96	0.97	1.00	0.37
Income	0.59	1.00	0.68	0.98	1.00	0.53	1.00
Assets and &Dwell. Ind.	0.96	1.00	1.00	1.00	1.00	0.96	1.00
Parental education							
Completed Kindergarden	0.90	1.00	0.95	1.00	1.00	1.00	1.00
Completed Secondary	0.99	1.00	0.68	0.96	1.00	0.99	0.76
Parental aspirations							
Age to stop education	0.99	1.00	1.00	1.00	1.00	1.00	1.00
Expects child to complete secondary	0.80	1.00	0.95	1.00	0.81	1.00	0.89
Percieved discrimination							
Household percievs disc. Toward RT at school (Ind.)	1.00	1.00	0.79	0.98	0.96	0.98	0.94
Community percievs disc. Toward RT at school (Ind.)	0.80	0.99	1.00	0.98	1.00	1.00	1.00
Treated children	547	625	619	641	620	579	601
Control children	631	631	631	631	631	631	631
Total observations	4863	4863	4863	4863	4863	4863	4863

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.

Table 5: 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	1.40	1.18
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	1.40	1.18
Observations	3,156	1,707	3,156	1,707	3,156	1,707
Equality of all treatment arms						
F-statistic	5.48	3.60	4.98	3.89	2.30	9.96
p-value	0.48	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: Treatment effects on child development

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	0.0280	-0.111	-0.0531	-0.134*	0.0268	-0.111	0.151**	-0.0222
	(0.0491)	(0.0787)	(0.0438)	(0.0700)	(0.0523)	(0.0831)	(0.0718)	(0.0814)
Control group mean of dep. var.	1.017	0.647	1.536	1.120	1.426	1.125	1.426	1.089
Observations	3,116	1,701	3,116	1,701	3,116	1,701	3,116	1,701

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	0.0288	0.0133	-0.0954	-0.0837	-0.00488	-0.0302	0.120	-0.00277
	(0.0723)	(0.0875)	(0.0750)	(0.0856)	(0.0736)	(0.0914)	(0.0844)	(0.0946)
Free access	-0.0417	-0.147	-0.175***	-0.166*	-0.0472	-0.0722	0.112	-0.0530
	(0.0636)	(0.0961)	(0.0623)	(0.0880)	(0.0676)	(0.0981)	(0.0839)	(0.114)
Fre access+ Info.	0.0556	-0.162	-0.0432	-0.164*	0.0495	-0.178*	0.184**	-0.0838
	(0.0627)	(0.106)	(0.0625)	(0.0910)	(0.0703)	(0.0999)	(0.0897)	(0.117)
7 lev	0.0595	-0.158*	0.0536	-0.135	0.0694	-0.170	0.139	0.00675
	(0.0738)	(0.0895)	(0.0636)	(0.0992)	(0.0813)	(0.104)	(0.0850)	(0.103)
7 lev + Info	0.0675	-0.108	-0.0271	-0.0799	0.0543	-0.111	0.174*	0.0413
	(0.0690)	(0.0984)	(0.0521)	(0.0799)	(0.0651)	(0.0932)	(0.0960)	(0.0985)
20 lev	0.0260	-0.0817	-0.0230	-0.127	0.0526	-0.0426	0.112	-0.0198
	(0.0584)	(0.0993)	(0.0522)	(0.0804)	(0.0640)	(0.101)	(0.0873)	(0.0999)
20 lev + Info	0.00638	-0.112	-0.0577	-0.130	0.0136	-0.156	0.209**	-0.0420
	(0.0614)	(0.108)	(0.0469)	(0.0892)	(0.0603)	(0.107)	(0.0885)	(0.111)
Control group mean of dep. var.	1.017	0.647	1.536	1.120	1.426	1.125	1.426	1.089
Observations	3,116	1,701	3,116	1,701	3,116	1,701	3,116	1,701
Equality of all treatment arms								
F-statistic	3.993	8.170	13.07	2.481	4.406	6.529	3.062	1.847
p-value	0.678	0.226	0.0420	0.871	0.622	0.367	0.801	0.933

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.

Table 7: Production function of child development: substitution and intensity parameters, marginal products

Dependent variable	Literacy (1)	Numeracy (2)	Socioemotional (3)	Motor (4)
Substitution parameter				
ρ	0.49	0.5	0.42	0.45
	[0]	[0]	[0]	[0]
Factor intesity				
β	0.5	0.56	0.57	0.47
	[0]	[0]	[0]	[0]
Median marginal product				
School attendance	0.25	0.49	0.46	0.24
	[0]	[0]	[0]	[0]
Parental investment	0.23	0.35	0.31	0.24
	[0]	[0]	[0]	[0]
Control functions				
ν_{Sc} (Parental investment)	-0.36	-0.35	-0.37	-0.41
	[0]	[0]	[0]	[0]
First stage F-statistic	[29.7]	[29.7]	[29.7]	[29.7]
ν_{Pi} (School attendance)	-0.4	-0.46	-0.51	-0.37
	[0]	[0]	[0]	[0]
First stage F-statistic	[39.3]	[39.3]	[39.3]	[39.3]
Observations	4,863	4,863	4,863	4,863

Table 8: Production function of child development: Productivity shifters

Dependent variable	Literacy (1)	Numeracy (2)	Socioemotional (3)	Motor (4)
Productivity parameters				
A (Constant)	-0.95 [0]	-0.25 [0]	-0.04 [0.69]	-0.94 [0]
A_f (Female FE)	-0.07 [0.46]	-0.13 [0.1]	-0.08 [0.23]	0.1 [0.18]
A_{A4} (Age=4 FE)	0.19 [0]	0.2 [0]	0.12 [0.08]	0.23 [0]
A_{A5} (Age=5 FE)	0.57 [0]	0.52 [0]	0.37 [0]	0.62 [0]
A_{A6} (Age=6 FE)	0.96 [0]	0.77 [0]	0.6 [0]	1.01 [0]
$A_{f,A4}$ (Age=4#Female FE)	0.1 [0.22]	0.16 [0.04]	0.09 [0.21]	0.03 [0.74]
$A_{f,A5}$ (Age=5#Female FE)	0.12 [0.18]	0.16 [0.04]	0.11 [0.06]	0.09 [0.32]
$A_{f,A6}$ (Age=6#Female FE)	0.08 [0.34]	0.1 [0.21]	0.03 [0.62]	-0.02 [0.86]
A_{RT} (Roma or Turkish FE)	-0.26 [0]	-0.15 [0.01]	-0.16 [0.01]	-0.22 [0]
A_{PI} (Enrolled at Baseline)	-0.05 [0.09]	-0.09 [0.01]	-0.15 [0]	-0.08 [0.03]
A_{PI} (Baseline Parent. Inv.)	-0.03 [0.14]	-0.04 [0.02]	-0.07 [0]	-0.07 [0]
A_{PEd} (Parents completed sec.)	0.13 [0]	0 [0.94]	-0.01 [0.61]	0.06 [0.05]
A_{PEx} (Parents expect sec. ed.)	0.09 [0.04]	-0.06 [0.12]	-0.09 [0.01]	0.08 [0.11]
A_{Disc} (Discrimination)	0.12 [0.07]	0.17 [0]	0.16 [0.02]	0.05 [0.56]
$A_{Disc\#RT}$ (Discrimination # RT)	0.11 [0.15]	0.06 [0.27]	0.09 [0.16]	0.24 [0.03]
Observations	4,863	4,863	4,863	4,863

Table 9

	Observed difference			
Developmental outcomes	Total	School attendance	P. Investment	TFP
Literacy				
Mean	-0.40	-0.07	-0.12	-0.20
Explained %		[17.5%]	[30%]	[50%]
Numeracy				
Mean	-0.25	-0.06	-0.13	-0.05
Explained %		[24%]	[52%]	[20%]
Socioemotional				
Mean	-0.21	-0.06	-0.14	0.00
Explained %		[28.6%]	[66.7%]	[0%]
Motor				
Explained %	-0.21	-0.07	-0.11	-0.01
		[33.3%]	[52.4%]	[4.8%]

Table 10: Production Function first stage

Variables	(1) Parental investment	(2) School attendance
Minority	-0.04	0.27**
	(0.09)	(0.11)
Enrolled at baseline	0.04	0.58***
	(0.03)	(0.04)
Baseline PI	0.23***	-0.00
	(0.03)	(0.03)
Female	0.14**	0.08
	(0.07)	(0.11)
Age=4	0.12*	0.04
	(0.07)	(0.08)
Age=5	0.10	-0.04
	(0.07)	(0.08)
Age=6	0.10	-0.13
	(0.07)	(0.08)
Female-Age=4	-0.12	-0.09
	(0.08)	(0.12)
Female-Age=5	-0.12	-0.08
	(0.08)	(0.12)
Female-Age=6	-0.14*	-0.08
	(0.08)	(0.13)
Parents expect secondary ed.	0.29***	0.08
	(0.04)	(0.06)
H. Head completed secondary ed.	0.21***	0.14***
	(0.04)	(0.04)
Com. percieves disc.	-0.13	0.04
	(0.09)	(0.12)
Minority#Com.percieves disc.	-0.04	-0.30**
	(0.10)	(0.13)
Parental Investment IV	0.83***	-0.12
	(0.15)	(0.18)
Attendance IV	-0.09	0.91***
	(0.13)	(0.15)
Constant	-0.24	-0.21
	(0.25)	(0.28)
R-squared	0.18	0.15
Parental Inv. F-stat	29.7	
Attendance F-stat		39.3
Observations	4,863	4,863

Appendix

A Additional Tables

Table A1: Content of the Information Intervention

Info session	Short description	Length	Time period of implementation	Leading the session	Attendees
1	Data on benefits of early learning, description of the project and logistics of enrolment in kindergarten.	60 minutes	July- beginning of September 2014	NGO representative	Parents from the whole community
2	Teacher's presentation on the setting and daily regime, curriculum and goals. NGO repeats logistics of enrolment in kindergarten.	50 minutes	September-October 2014	KG teacher and NGO representative	Parents from the whole community
3	Open house. Observation at the kindergarten of a celebration or a lesson open to the parents.	45 minutes	December 2014, March 2015 or May 2015	KG teacher and NGO representative – observing the children perform	Parents from the whole community
4	Teacher's presentation on children's progress. NGO training on parent-teacher-child effective communication.	70 minutes	February-March 2015	KG teacher and NGO representative	Parents from the whole community
5	Role model: community representative's presentation on importance of kindergarten for success in school and later on in life.	90 minutes	May-June 2015	NGO representative and Roma role-model	Parents from the whole community

Table A2: Treatment effect on parental investment components

Panel A: Pooled treatment effect

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable	Read		Tell stories		Sing		Play		Name	
Enrolled at baseline	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Treated	-0.04	-0.11	-0.06	-0.21	-0.11	-0.47***	-0.21**	-0.37**	-0.05	-0.25
	(0.10)	(0.19)	(0.09)	(0.14)	(0.11)	(0.12)	(0.11)	(0.15)	(0.11)	(0.16)
Control group mean of dep. var.	0.050	-0.23	0.018	-0.28	-0.11	-0.27	-0.11	-0.28	0.0058	-0.22
Observations	3,156	1,707	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)	(9)	(10)
Dependent variable	Read		Tell stories		Sing		Play		Name	
Enrolled at baseline	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Information	0.06	-0.17	0.04	-0.17	-0.15	-0.47**	-0.02	-0.18	0.01	-0.15
	(0.15)	(0.23)	(0.15)	(0.19)	(0.17)	(0.22)	(0.14)	(0.20)	(0.15)	(0.23)
Free access	-0.14	-0.36*	-0.19	-0.49***	-0.23	-0.70***	-0.19	-0.41**	-0.16	-0.47**
	(0.15)	(0.19)	(0.14)	(0.16)	(0.16)	(0.17)	(0.14)	(0.18)	(0.15)	(0.19)
Fre access+ Info.	0.01	-0.11	-0.04	-0.35**	-0.09	-0.56***	-0.21	-0.32	0.01	-0.33
	(0.15)	(0.22)	(0.12)	(0.17)	(0.15)	(0.20)	(0.16)	(0.21)	(0.15)	(0.20)
7 lev	-0.01	-0.14	-0.05	-0.25	-0.00	-0.42***	-0.34**	-0.46**	0.09	-0.21
	(0.13)	(0.20)	(0.13)	(0.16)	(0.13)	(0.14)	(0.15)	(0.19)	(0.14)	(0.19)
7 lev + Info	-0.04	0.18	0.00	-0.04	-0.17	-0.39*	-0.17	-0.17	-0.01	-0.11
	(0.13)	(0.24)	(0.12)	(0.20)	(0.15)	(0.20)	(0.15)	(0.20)	(0.15)	(0.23)
20 lev	-0.07	0.01	-0.09	-0.10	-0.17	-0.37**	-0.22	-0.44**	-0.17	-0.15
	(0.14)	(0.21)	(0.12)	(0.17)	(0.15)	(0.16)	(0.14)	(0.21)	(0.15)	(0.21)
20 lev + Info	-0.05	-0.19	-0.07	-0.16	0.02	-0.45**	-0.30*	-0.63***	-0.06	-0.36*
	(0.14)	(0.21)	(0.14)	(0.17)	(0.14)	(0.20)	(0.16)	(0.21)	(0.16)	(0.19)
Control group mean of dep. var.	0.050	-0.23	0.018	-0.28	-0.11	-0.27	-0.11	-0.28	0.0058	-0.22
Observations	3,156	1,707	3,156	1,707	3,156	1,707	3,156	1,707	3,156	1,707
Equality of all treatment arms										
F-statistic	1.88	16	2.64	13.4	4.27	4.18	5.93	6.51	5.03	6.22
p-value	0.93	0.014	0.85	0.037	0.64	0.65	0.43	0.37	0.54	0.40

Table A3: Treatment effect on income.

Panel A: Pooled treatment effect

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Income (%)		Labor income (%)		Unemployed	
Enrolled at baseline	Yes	No	Yes	No	Yes	No
Treated	-0.02	0.02	-0.08	-0.04	-0.00	-0.01***
	(0.07)	(0.10)	(0.07)	(0.11)	(0.00)	(0.00)
Control group mean of dep. var.	0.19	0.16	0.24	0.16	0.37	0.40
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	Income (%)		Labor income (%)		Unemployed	
Enrolled at baseline	Yes	No	Yes	No	Yes	No
Information	-0.00	0.05	-0.17*	-0.09	-0.02	-0.01
	(0.09)	(0.13)	(0.09)	(0.13)	(0.01)	(0.01)
Free access	-0.02	-0.02	-0.03	0.03	-0.00	-0.00
	(0.10)	(0.12)	(0.13)	(0.15)	(0.00)	(0.00)
Fre access+ Info.	-0.04	0.03	-0.16	-0.02	0.00	-0.00
	(0.08)	(0.11)	(0.10)	(0.15)	(0.00)	(0.00)
7 lev	-0.05	0.01	-0.11	-0.07	-0.01	-0.02**
	(0.10)	(0.13)	(0.11)	(0.15)	(0.01)	(0.01)
7 lev + Info	-0.01	0.14	-0.12	0.02	-0.01	-0.01
	(0.09)	(0.11)	(0.12)	(0.14)	(0.01)	(0.01)
20 lev	-0.02	0.03	-0.00	-0.08	0.00	-0.01
	(0.09)	(0.15)	(0.10)	(0.16)	(0.00)	(0.01)
20 lev + Info	-0.02	-0.08	-0.01	-0.09	0.00	-0.01
	(0.08)	(0.11)	(0.09)	(0.12)	(0.00)	(0.01)
Control group mean of dep. var.	0.19	0.16	0.24	0.16	0.37	0.40
Observations	3,156	1,707	3,156	1,707	3,156	1,707
Equality of all treatment arms						
F-statistic	0.34	7.56	5.70	1.56	10.3	10.1
p-value	1	0.27	0.46	0.96	0.11	0.12

Table A4: Treatment effect on KG perceptions (Part 1).

Panel A: Pooled treatment effect

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable	School distance		Tuition		Lunch fee		Transport fee	
Enrolled at baseline	Yes	No	Yes	No	Yes	No	Yes	No
Treated	0.03	0.02	0.55	0.10	-0.01	0.33	-0.38	-0.08
	(0.11)	(0.03)	(0.47)	(0.31)	(0.10)	(0.24)	(0.38)	(0.31)
Control group mean of dep. var.	2.36	3.98	28.9	26.8	1.73	-0.98	1.51	0.47
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	School distance		Tuition		Lunch fee		Transport fee	
Enrolled at baseline	Yes	No	Yes	No	Yes	No	Yes	No
Information	-0.02	0.00	0.55	-0.16	0.17	0.45*	-0.33	-0.01
	(0.13)	(0.03)	(0.47)	(0.40)	(0.15)	(0.28)	(0.51)	(0.37)
Free access	0.10	0.02	0.69	0.49	0.09	0.38	0.03	0.21
	(0.13)	(0.02)	(0.50)	(0.36)	(0.13)	(0.27)	(0.57)	(0.37)
Fre access+ Info.	-0.06	0.01	0.54	-0.31	0.06	0.04	-0.94**	0.04
	(0.14)	(0.03)	(0.51)	(0.38)	(0.13)	(0.28)	(0.46)	(0.36)
7 lev	0.08	0.02	0.55	0.20	-0.02	0.52**	0.00	-0.09
	(0.17)	(0.02)	(0.56)	(0.36)	(0.13)	(0.26)	(0.57)	(0.35)
7 lev + Info	0.16	0.03	0.62	-0.04	0.00	0.17	-0.04	-0.10
	(0.13)	(0.03)	(0.60)	(0.40)	(0.12)	(0.29)	(0.53)	(0.38)
20 lev	-0.00	0.02	0.15	0.66	-0.10	0.59*	-0.58	-0.25
	(0.14)	(0.03)	(0.58)	(0.42)	(0.19)	(0.35)	(0.51)	(0.43)
20 lev + Info	-0.03	0.01	0.79	0.08	-0.24	0.03	-0.76	-0.53
	(0.12)	(0.03)	(0.49)	(0.36)	(0.17)	(0.34)	(0.48)	(0.40)
Control group mean of dep. var.	2.36	3.98	28.9	26.8	1.73	-0.98	1.51	0.47
Observations	3,156	1,707	3,156	1,707	3,156	1,707	3,156	1,707
Equality of all treatment arms								
F-statistic	5.23	8.96	2.96	11	6.18	10.7	6.88	4.80
p-value	0.52	0.18	0.81	0.087	0.40	0.099	0.33	0.57

Table A5: Treatment effect on KG perceptions (Part 2).

Panel A: Pooled treatment effect

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent variable	School etihnic comp.		RT respected by staff		RT respected by children		RT respected by teachers		KG important for PE		KG important for SE	
Enrolled at baseline	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Treated	0.00	0.05	0.01	0.00	-0.02	0.12**	0.01	0.11*	0.01	0.09	0.04	0.13**
	(0.13)	(0.05)	(0.03)	(0.00)	(0.05)	(0.06)	(0.04)	(0.05)	(0.04)	(0.06)	(0.05)	(0.06)
Control group mean of dep. var.	1.85	3.97	3.76	4	3.67	3.61	3.68	3.61	3.67	3.61	3.63	3.57
Observations	3,156	1,707	3,156	1,707	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)	(9)	(10)	(11)	(12)
Dependent variable	School etihnic comp.		RT respected by staff		RT respected by children		RT respected by teachers		KG important for PE		KG important for SE	
Enrolled at baseline	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Information	0.12	0.06	-0.01	0.00	-0.02	0.19***	0.02	0.18***	-0.02	0.15**	0.01	0.20***
	(0.17)	(0.05)	(0.04)	(0.00)	(0.06)	(0.07)	(0.06)	(0.07)	(0.06)	(0.07)	(0.06)	(0.08)
Free access	0.26	0.06	-0.07*	0.00	-0.13**	0.06	-0.12*	0.05	-0.12*	0.01	-0.09	0.08
	(0.21)	(0.05)	(0.04)	(0.00)	(0.06)	(0.07)	(0.06)	(0.07)	(0.06)	(0.07)	(0.07)	(0.07)
Fre access+ Info.	-0.10	0.04	0.01	0.00	0.04	0.10	0.07	0.09	0.07	0.08	0.09	0.15
	(0.19)	(0.05)	(0.04)	(0.00)	(0.05)	(0.09)	(0.05)	(0.09)	(0.05)	(0.09)	(0.06)	(0.09)
7 lev	0.00	0.06	0.02	0.00	0.01	0.10	0.02	0.11	0.06	0.12*	0.07	0.14*
	(0.16)	(0.05)	(0.03)	(0.00)	(0.06)	(0.07)	(0.06)	(0.07)	(0.05)	(0.07)	(0.06)	(0.08)
7 lev + Info	-0.05	0.07	0.08**	-0.00	-0.01	0.12	0.01	0.09	-0.03	0.09	0.03	0.11
	(0.15)	(0.05)	(0.04)	(0.00)	(0.05)	(0.08)	(0.05)	(0.08)	(0.05)	(0.08)	(0.06)	(0.09)
20 lev	-0.06	0.04	0.00	0.01	-0.02	0.11	0.02	0.10	0.03	0.09	0.08	0.13
	(0.17)	(0.05)	(0.03)	(0.00)	(0.05)	(0.08)	(0.05)	(0.08)	(0.05)	(0.08)	(0.06)	(0.08)
20 lev + Info	-0.11	0.04	0.02	0.00	0.01	0.08	0.04	0.07	0.06	0.07	0.09	0.14*
	(0.16)	(0.05)	(0.04)	(0.00)	(0.06)	(0.07)	(0.05)	(0.06)	(0.05)	(0.07)	(0.06)	(0.07)
Control group mean of dep. var.	1.85	3.97	3.76	4	3.67	3.61	3.68	3.61	3.67	3.61	3.63	3.57
Observations	3,156	1,707	3,156	1,707	3,156	1,707	3,156	1,707	3,156	1,707	3,156	1,707
Equality of all treatment arms												
F-statistic	5.22	7.30	17.5	9.19	10.7	4.57	12.1	5.31	13.2	5.44	10.3	3.41
p-value	0.52	0.29	0.0077	0.16	0.097	0.60	0.059	0.50	0.040	0.49	0.11	0.76

B Model identification

B.1 Setup

As discussed in section 5, our structural model is characterized by the equation that describes the production function and two equations that describe how socio-demographic characteristics and the intervention affect preschool attendance and parental investment.

Taking logarithm of the equation that describes production yields:

$$ln\ \ y^d_{igap} = \alpha^d_{ag} + \gamma^d X_{ip} + \frac{1}{\rho^d}\Big(\beta^d \big(I^d_{ip}\big)^\rho + (1-\beta^d)\big(S^d_{ip}\big)\Big) + \eta^d_{ip} \tag{9}$$

Where $d \in \{Numeracy, Literacy, Motor, Socioemotional\}$ 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 an age-gender fixed effect α_{ag}, parent and child observable characteristics (X_{ip}) such as the education and expectations of parents, baseline inputs, ethnicity and perceived discrimination, and an unobserved component η^d_{ip}.

Parental investment and preschool attendance choices are made before the IDELA test is administered. This means that parents and children choose inputs before y^d_{igap} is realized. Thus, we assume that η^d_{ip} is only partially observed by parents. Formally, we decompose η^d_{ip} as follows:

$$\eta^d_{ip} = \nu_i + \nu_p + \varepsilon^d_{ip} \tag{10}$$

Where ν_i and ν_p represent heterogeneity in children and parents' productivity that is common across developmental outcomes and observed by parents before the IDELA test is administered but not by the econometrician. ε_{ip} accounts for the remaining heterogeneity that is unobserved by both parents and researchers at the time that investment and attendance choices are made. We assume that inputs are independent from ε_{ip}, this is consistent with the assumption that parents cannot systematically predict productivity shocks that have not been realized at the time inputs are chosen.

Finally, the equations that describe how socio-demographic characteristics and the intervention affect preschool attendance and parental investment are as follows:

$$I_{ip} = \gamma^I_{ag} + \kappa^I X_{ip} + g^I\big(D_{ip}\big) + \phi^I \nu_i + \psi^I \nu_p + \xi^I_{ip} \tag{11}$$

$$S_{ip} = \gamma^S_{ag} + \kappa^S X_{ip} + g^S\big(D_{ip}\big) + \phi^S \nu_i + \psi^S \nu_p + \xi^S_{ip} \tag{12}$$

For $L = I, S$ the term γ^L_{ag} represents a gender-age fixed effect, we also include the remaining determinants of productivity from equation, namely X_{ip} which comprises observed determinants plus ν_i and ν_p which account for the unobserved determinants that parents observe. ξ^L_{ip} accounts for remaining unobserved preferences for inputs which we assume to be uncorrelated with age, gender, X_{ip}, ν_i and ν_p.

We replace $g^L(D_{ip})$ with the conditional expectation of I and S given age, gender and the treatment (which can be non-parametrically identified using saturated regression). Thus, equations 11 and 12 become:

$$I_{ip} = \gamma^I_{ag} + \kappa^I X_{ip} + \lambda^I \mu^I_{a,g,t} + \phi^I \nu_i + \psi^I \nu_p + \xi^I_{ip} \tag{13}$$

$$S_{ip} = \gamma^S_{ag} + \kappa^S X_{ip} + \lambda^S \mu^S_{a,g,t} + \phi^S \nu_i + \psi^S \nu_p + \xi^S_{ip} \tag{14}$$

B.2 Identification

We assume that ν_i, ν_p, ξ^I_{ip} and ξ^S_{ip} are independent from socio-demographic characteristics (the components of X_{ip} plus age and gender) and the intervention. This means that γ^L_{ag}, κ^L and λ^L from equations 13 and 14 can be identified using linear regression. From here, the terms $e^L_i \equiv \phi^I \nu_i + \psi^I \nu_p + \xi^I_{ip}$ are also identified for each child i.

Now consider a subpopulation of children such that $I_{ip} = I^*$ and $S_{ip} = S^*$ for some I^* and S^* in the support of I and S. Let a^* and g^* be the age and gender for the excluded category from the saturated regression used to identify equations 13 and 14. Now, note that:

$$\mathbb{E}[\Delta ln\ \ y^d_{igap}|I = I^*, S = S^*] - \mathbb{E}[\Delta ln\ \ y^d_{igap}|I = I^*, S = S^*, a = a^*, g = g^*] \tag{15}$$

$$= \alpha^d_{ag} + \gamma^d\Big(\mathbb{E}[X_{ip}|I = I^*, S = S^*] - \mathbb{E}[X_{ip}|I = I^*, S = S^*, a = a^*, g = g^*]\Big) \tag{16}$$

$$+ \Big(\mathbb{E}[\nu_i + \nu_p|I = I^*, S = S^*] - \mathbb{E}[\nu_i + \nu_p|I = I^*, S = S^*, a = a^*, g = g^*]\Big) \tag{17}$$

$$+ \Big(\mathbb{E}[\varepsilon^d_{ip}|I = I^*, S = S^*] - \mathbb{E}[\varepsilon^d_{ip}|I = I^*, S = S^*, a = a^*, g = g^*]\Big) \tag{18}$$

Where the last term is equal to zero because ε^d_{ip} is assumed to be independent from age, gender and inputs.

Note that:

$$\mathbb{E}[\Delta(e^L_i)] \equiv \mathbb{E}[e^L_i|I = I^*, S = S^*] - \mathbb{E}[e^L_i|I = I^*, S = S^*, a = a^*, g = g^*] \tag{19}$$

$$= \phi^L\Big(\mathbb{E}[\nu^L_i|I = I^*, S = S^*] - \mathbb{E}[\nu_i|I = I^*, S = S^*, a = a^*, g = g^*]\Big) \tag{20}$$

$$+ \psi^L\Big(\mathbb{E}[\nu^L_p|I = I^*, S = S^*] - \mathbb{E}[\nu_p|I = I^*, S = S^*, a = a^*, g = g^*]\Big) \tag{21}$$

Moreover, if ϕ and ψ are different from zero and from each other, a unique linear combination of e^I_i and e^S_i exists such that:

$$k^I\mathbb{E}[\Delta(e^L_i)] + k^S\mathbb{E}[\Delta(e^L_i)] = \Big(\mathbb{E}[\nu_i + \nu_p|I = I^*, S = S^*] - \mathbb{E}[\nu_i + \nu_p|I = I^*, S = S^*, a = a^*, g = g^*]\Big) \tag{22}$$

And therefore:

$$k^I\mathbb{E}[e^L_i] + k^S\mathbb{E}[e^L_i] = \mathbb{E}[\nu_i + \nu_p] \tag{23}$$

Furthermore, $\Delta(e^I_i)$ and $\Delta(e^S_i)$ are linearly independent from X and the age-gender indicators since they belong to the space generated by treatment assignment which is random and therefore we assume is linearly independent from any other observable characteristic.

Therefore, the conditional expectation from equation 15 is a unique linear combination of the age and gender indicators, the difference of expectations given by $\Big(\mathbb{E}[X_{ip}|I = I^*, S = S^*] - \mathbb{E}[X_{ip}|I = I^*, S = S^*, a = a^*, g = g^*]\Big)$, $\Delta(e^I_i)$ and $\Delta(e^S_i)$. This means that linear regression of $\Delta ln\ \ y^d_{igap} - \mathbb{E}[\Delta ln\ \ y^d_{igap}|I = I^*, S = S^*, a = a^*, g = g^*]$ on the age-gender indicators, X, $\Delta(e^I_i)$ and $\Delta(e^I_i)$ restricted to the subpopulation with $I = I*$ and $S = S^*$ identifies α^d_{ag}, γ^d, k^I and k^S. From here, $\mathbb{E}[\nu_i + \nu_p|X, a, g, I, S]$ is also identified.

Finally, taking expectation of equation 9 yields:

$$\mathbb{E}[ln\ \ y^d_{igap}|S, I] = \alpha^d_{ag} + \gamma^d\mathbb{E}[X_{ip}|S, I] + \Big[\frac{1}{\rho^d}\Big(\beta^d\big(I^d_{ip}\big)^\rho + (1 - \beta^d)\big(S^d_{ip}\big)\Big)\Big] + \mathbb{E}[\nu_i + \nu_p|S, I] \tag{24}$$

Equation 24 is non-linear in ρ and β and we cannot prove that it is guaranteed to have a solution or that, when a solution exists, it is unique. However, during estimation, we tested several initial values of ρ in the interval [-100,1) and β in the interval (0,1) and found that in all cases our estimation procedure converged to the same solution. This suggests that in our sample and population equation 24 might have a unique solution and therefore is identified.