The impact of early childhood parenting practices on skills development. Evidence from Hungary

by

Sergiu Burlacu

Submitted to

Central European University

Department of Economics

In partial fulfilment of the requirements for the degree of Master of Arts

Supervisor: Professor Gabor Kezdi

Budapest, Hungary

2016

Abstract

Using a rich longitudinal dataset following 10,000 students, this thesis investigates the impact of early childhood parental investment on socio-emotional and cognitive skills, and health during adolescence and early adulthood. Early childhood is a very salient development stage with very high potential returns on investment. However, the mechanisms behind are still not very well understood: how skills interact, how parents make investment decisions or what is the optimal timing and type of intervention to promote the growth of a certain skill. I estimate OLS and IV models where I control for socio-economic status and early childhood health. Early childhood investment is instrumented through birth order and whether a parent/stepparent left/joined the household during early childhood. The IV strategy indicates a significant downward bias in the OLS model suggesting that parents compensate for unobserved shocks in the development of the child. The estimates indicate that early childhood investment is strongly associated with socio-emotional skills and academic achievement, but less with mathematics test scores and health. Furthermore, I find some evidence that time investments matter more for socio-emotional skills while material investment are more important for cognitive skills, and that early health strongly predicts skills during adolescence. Finally, I find that after controlling for socio-economic status and early childhood health, there is no difference in socio-emotional skills or parental investment between Roma and non-Roma students.

KEYWORDS: child development, early childhood parenting, socio-emotional skills, Romani

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Professor Gabor Kezdi for his ongoing support and insightful feedback, and my sister Diana for always being a role model, mentor and best friend. I would also like to thank my wonderful family and friends, my CEU classmates Gabor, Robert and Zsolt for helping me with translations of the questionnaire in Hungarian when needed, the Roma Graduate Preparation Program for providing me the great experience of tutoring, and to the entire CEU community for being such an important part of my life.

Table of Contents

1. Introduction
2. Literature Review
3. Theoretical Framework
4. Data and Variables
4.1. Data
4.2. Dependent Variables
4.3. Explanatory Variables
5. Empirical Model
6. Regression Results
6.1. OLS regressions
6.2. Instrumental variable regressions
7. Conclusion
References
Appendix

1. Introduction

While the production function of human capital has long interested labor economists, early research did not acknowledge the fact that various periods in childhood are more sensitive than others to the optimal development of the individual (Becker and Tomes, 1986). The literature on early childhood development has seen a substantial growth in the past fifteen years with the increasing realization that early life conditions are strong predictors of adults' outcomes (Almond and Currie, 2011; Walker et al., 2007; Anderson, 2008; Currie and Thomas, 1999). Early childhood is a salient development period for future outcomes as far as employment opportunities, wages, and adult health¹.

Human capital development is a very dynamic process with multiple dimensions (health, cognition and socio-emotional skills) interacting at different time periods and with various inputs. Past levels of human capital influence its growth later in life, making the identification of particularly salient time periods very important. Attanasio (2015) notes that there still are many elements of the human capital development process which are not very well understood. Less is known regarding which years are more salient for certain skills, what mechanisms generate the development gaps, how certain skills interact with one another to foster or hinder growth, what type of investments matter for what skills and how parents choose how much to invest in their children.

Early childhood seems to be a crucial development stage providing an important window of opportunity for intervention. Heckman et al. (2010) stresses out that children are more malleable during this time period, making them also more vulnerable to shocks. They also point out that investments during these years show very high returns and could help reduce schooling and labor market opportunity gaps. However, even early childhood is not a homogenous life

¹ Almond and Currie (2011), Engle et al. (2011) and Walker et al. (2011) provide comprehensive reviews of early childhood evidence from developed and developing countries.

stage. Recent research shows that to achieve long lasting impact on cognitive skills, interventions might be more effective if they are implemented in the first 3 years of life, making this period crucial for optimal cognitive skills development (Heckman et al., 2013). Investment around preschool age with long term impact usually did not have a lasting effect on cognitive skills but improved socio-emotional skills which seem to be a crucial skill for academic achievement and labor market outcomes (Attanasio, 2015; Heckman et al., 2013) Understanding all these issues is extremely important in order to design effective interventions that stimulate human capital development and increase productivity.

This thesis aims to investigate the long-term link between early childhood parental investment and the development of socio-emotional skills, cognitive skills and health during adolescence and early adulthood, using a rich and unique longitudinal dataset following 10,000 students for more than 6 years in Hungary. While causality is hard to establish, the model controls for household's socio-economic status, early childhood health, home environment and school fixed effects, and attempts to account for the endogeneity of investment through instrumental variable estimation. Parental investment is instrumented through birth order in early childhood and whether one of the parents stopped living in the household, or whether a stepparents joined the household during early childhood. I find large and significant effects on socio-emotional skills and academic achievement. The effects on cognitive skills and health are more ambiguous. Furthermore, following a new line of research opened by Attanasio et al. (2015), I find that time investments matter more for socio-emotional skills while material investments are more important for cognitive skills.

Secondly, I try to provide insights into the parental decision process regarding investment levels. Parents might choose to compensate for various shocks or differences (e.g. invest more in a child which had health issues in the first years of life) or to reinforce them (e.g. invest more in the child displaying higher cognitive or socio-emotional skills). Parents seem to respond differently to different shocks in different settings (Almond and Mazumder, 2013). In line with Cunha et al. (2010), Attanasio, Meghir and Nix (2015) and Attanasio et al. (2015), I find that parents compensate for negative shocks in early childhood through higher investments.

In addition, I find that after controlling for early childhood health, home environment, and school fixed effects, there is no association between early childhood investment and socioeconomic status. Poor parents are just as capable of providing optimal stimulations to their children. This raises an important research question: what are the factors (lack of information, lack of resources etc.) which stop poorer parents from investing more in their children?

Thirdly, benefiting from detailed ethnical information in the dataset and following Kertesi and Kezdi (2011, 2014), I investigate whether there is a gap between Roma and non-Roma students in terms of socio-emotional skills. While Roma students are less satisfied with their lives and have lower expectations for the future during early adulthood (likely due to economic hardship and racial segregation), there is no difference in terms of socio-emotional skills between Roma and non-Roma students. Moreover, after controlling for socio-economic status and early childhood health, there is no difference in investment between Roma and non-Roma parents in the main sample².

This thesis proceeds as follows. The next section discusses the most relevant literature on the topic. Section 3 presents the theoretical model of human capital development behind the estimation. Section 4 describes the dataset, the key variable (and their estimation), the controls and the instruments used. Section 5 motivates the empirical strategy used to estimate the results presented in section 6. Finally, I conclude and suggest further areas of research.

² However, for the full sample, Roma students have 1.5 standard deviations lower levels of parental investment

2. Literature Review

This section briefly summarizes some of the most relevant evidence on the returns of early childhood investment from experimental and non-experimental studies. Almond and Currie (2011) provide a comprehensive summary of most of the recent evidence in the early childhood literature. While I focus more on the impact on socio-emotional skills, health and cognitive skills are considered as well, taking into account the complementariness between them. I also try to provide evidence from long term programs in both developed countries (e.g. Abecederian, Perry-Preschool) and developing countries (studies in Guatemala and Jamaica).

I focus more on interventions with long term follow-ups or with important implications for this thesis. I start with the Nurse-Family Partnership program evaluated using random assignment in Olds et al. (2010) and Howard and Brooks-Gunn (2009). The program provided home visits from nurses, to low-income young mothers, from pregnancy to the time the child is two years old. The program improved IQ and mother labor outcomes, and reduced internalizing behavior, substance abuse and criminal behavior as the children grew up. The effect on IQ faded in time, suggesting that socio-emotional skills might better explain the results.

Another influential program which improved many long term outcomes (academic achievement, criminal behavior and others) but did not have a long lasting effect on IQ is the HighScope Perry Preschool Program, which provided high quality preschool education and home-visits for two years, to high-risk children, especially black children. The program is evaluated in Heckman et al. (2010) and Heckman et al. (2013) which claim annual rates of return of 7-10%. Heckman et al. (2013) show that the outcomes are explained by sustained improvements in socio-emotional skills.

Programs with lasting impact on cognitive skills are usually implemented much earlier in life, before age 3. One influential program was the Abecedarian which targeted black children.

One cohort of children was followed starting from the first months of life through the third grade of school. The program was very complex and had various components: full day high-quality preschool, nutrition and medical care, parent-teacher interaction and school support. The effect on IQ was persistent into adulthood (Campbell et al., 2002), especially among girls (Anderson, 2008). The program also generated long term adult outcomes when the children were in their thirties (see Campbell et al., 2014). Two large scale preschool interventions, Head Start and the Chicago Child-Parent Center, received significant attention but their long-term effect were not evaluated using randomized assignment. The evaluations of Head Start show mixed results while the Child-Parents Center program show results consistent with the Perry-Preschool program. What seems to be an important conclusion is that involving parents in the program matters greatly (see Heckman et al., 2013).

While most of the evidence comes from programs implemented in developed countries, two of the most impressive long term studies are interventions in developing countries. The first program was a randomized nutrition intervention in Guatemala, which administered fortified nutrition drinks to children in randomly selected villages. The program has two interesting features. Firstly, the children were exposed to the treatment at various ages providing evidence on sensitive periods for intervention. Secondly, even the outcomes of the offsprings of the children are being observed. The program had significant long term impacts, stronger amongst the children who received the intervention earlier in life, and even showed an effect on the offsprings of the girls in the study (Mallucio et al., 2009; Behrman et al. 2009).

One other long term intervention is the influential Jamaican study (see Gertler et al., 2014). 129 1-2 year-olds stunted children were randomly assigned to receive psycho-social stimulation, nutritional supplements, or both, for two years. While the effect on cognitive skills and other outcomes on the group that received only nutritional supplements faded over time, in the other two groups it remained substantial into adulthood, and those children caught up in

terms of earnings with a matched non-stunted group followed in the study. The intervention improved parenting practices in the short run, the treatment groups catching up with the nonstunted control group, but the effect faded out over time.

Some of the most important conclusions from these interventions, underlined in Heckman et al. (2013) and Attanasio (2015), is that in order to obtain long-lasting effects on cognitive skills, the intervention has to be delivered very early in life. Socio-emotional skills seem to be more malleable, with long-lasting effects that strongly impact future positive outcomes. While the interventions cited above had impressive outcomes, much is still unknown about what makes an intervention successful.

Attanasio (2015) emphasize what we do not know, which is the optimal timing, duration and intensity of interventions. Moreover, little is known about the roles of genes in generating or remediating development gaps. We still need to understand how parents, individually and as a household, makes decisions regarding investments, how they react to interventions, how they respond to shocks in their children's development and how they allocate scarce resources among siblings. Finally, we need to understand the complementarities between skills. Do emotional skills favor the development of cognitive skills or is it the other way round? Which skills are most sensitive for the development of future skills? Which skills should be addressed more intensively by interventions and at what ages?

3. Theoretical Framework

This section deals with the theoretical framework behind the estimation of the human capital production function, drawing on the work of Cunha and Heckman (2008), Cunha et al. (2010) and Attanasio et al. (2015). Todd and Wolpin (2003) propose a modelling framework for estimating the production function of cognitive skills. They emphasize the need to incorporate theoretical frameworks to unravel the mechanisms behind experiments and argue that while experiments are important for understanding the effect of an intervention or policy, they are not sufficient to elucidate how the intervention worked. Cunha et al. (2010), Attanasio, Meghir, and Nix (2015), Attanasio et al. (2015) and Attanasio (2015) formulate and estimate richer models of human capital development using more dimensions of human capital, such as health, cognitive and socio-emotional skills.

Human capital is viewed as a multidimensional model that evolves over time from the moment of conception. The literature usually considers health, cognition and socio-emotional skills as dimensions or factors of human capital. These factors evolve over time depending on their past levels and other environmental factors. As noted in Attanasio (2015), these environmental factors are either endogenous (parental investment) or can be considered exogenous to the evolution of human capital (e.g. family background). The first set of environmental factors raise more research questions because they are chosen by parents thus making them more prone to interventions.

Heckman (2007) emphasizes important features of the production function: selfproductivity, cross-productivity and dynamic complementarity. Self-productivity and crossproductivity imply that past levels of a skill promote the growth of that skill (self-productivity) or other skills (cross-productivity)³. The presence of dynamic complementarities means that

³ For example, a healthy child might end up spending more time with other children developing social skills. Or, as suggested by Heckman (2007) emotionally secure children might be more open to learn cognitive skills or take better care of their health.

past levels of skills make investments in those skills more productive later in life. Cunha and Heckman (2007) suggest that dynamic complementarity and self-productivity explain why early childhood programs are much more effective than teenage programs.

As mentioned above, parents choose how much to invest in their children, thus making the investment decision endogenous in the model. One important issue in the literature is how parents respond to certain shocks in the development of the child. Parents can choose to compensate or reinforce these shocks. Understanding the decision process of the parents is crucial for the identification of the effect of parental investment on child development. I discuss these issues further in the following sections. As presented in Attanasio (2015), when choosing the investment level, parents are assumed to maximize some objective function depending on the child's capabilities and household consumption. Many things might shape their investment decision such as beliefs, preferences across children, information, resources, financial constraints (credit, insurance etc.).

Adapting the model discussed in Attanasio (2015) and estimated in Attanasio et al. (2015), I assume that parents solve the following maximization problem:

$$Max_{\{C_{i,t},X_{i,t}\}}U(C_{i,t},H_{i,t+1})$$

$$\tag{1}$$

s. t.
$$C_{i,t} + P_t^x X_{i,t} = Y_{i,t}$$
 (2)

where
$$H_{i,t+1} = g_t(H_{i,t}, X_{i,t}, Z_{i,t}, e_{i,t})$$
 (3)

Parents choose an initial investment level $(X_{i,t})$ in order to maximize utility, which depends on initial consumption $(C_{i,t})$ and on the human capital level of the child in the future $(H_{i,t+1})$. They face a budget constraint that depends also on the price of investment (P_t^x) and available resources $(Y_{i,t})$. Future level of human capital is assumed to be a function of initial level of human capital $(H_{i,t})$, initial investment, environmental factors $(Z_{i,t})$ and shocks $(e_{i,t})$. Human capital, investments and environmental factors are assumed to be multidimensional objects:

$$H_{i,t} = \left\{ \theta_{i,t}^C, \theta_{i,t}^S, \theta_{i,t}^H \right\}$$
(4)

$$X_{i,t} = \left\{ \theta_{i,t}^M, \theta_{i,t}^T \right\}$$
(5)

$$Z_{i,t} = \left\{ Z_{i,t}^{M}, Z_{i,t}^{F}, Z_{i,t}^{R} \right\}$$
(6)

Human capital has cognitive $(\theta_{i,t}^{C})$, socio-emotional $(\theta_{i,t}^{S})$ and health $(\theta_{i,t}^{H})$ dimensions; investment has material $(\theta_{i,t}^{M})$ and time $(\theta_{i,t}^{T})$ dimensions; while the environmental factors are divided in mother background $(Z_{i,t}^{M})$, father background $(Z_{i,t}^{F})$ and other environmental factors $(Z_{i,t}^{R})$. Finally, the investment variable depends on initial level of capital (reinforcing or compensating behavior), background variables, resources, prices, shocks to investment $(e_{i,t}^{x})$ and preferences parameters (π) :

$$X_{i,t} = f(H_{i,t}, Z_{i,t}, P_t^{\chi}, Y_{i,t}, e_{i,t}^{\chi}, \pi)$$
(7)

Attanasio (2015) identifies two main issues in this framework. Firstly, the investment variable is endogenous in the production function since shocks to child development are likely to be correlated with investment decisions - $E[X_{i,t} | e_{i,t}^x] \neq 0$ (reinforcing or compensating behavior). This issue is dealt with by using the exogenous variation in P_t^x through measures of wages and labor market conditions. Secondly, the variables in the models are not directly observed. Attanasio et al. (2015) apply the conceptual framework from Cuhna et al. (2010) and construct a measurement system for each latent variable and estimates them using exploratory factor analysis.

Attanasio et al. (2015) estimate the parameters using a CES production function specification for equation (3) and find that the production function is approximated by a Cobb Douglas shape. They find strong evidence of self-productivity of both cognitive and emotional skills but only find evidence of cross-productivity on emotional skills with respect to cognitive skills. This last result is contrary to Cunha et al. (2010) which find that, for a sample of older children and in a higher income country, initial socio-emotional skills are important for future cognition, but not the other way round. Furthermore, material investment matters more for cognitive skills while time invested is more important for emotional skills (Attanasio, 2015).

(3) discussed in the section 5 and the potential limitations of the analysis.

4. Data and Variables

4.1. Data

This thesis uses the Hungarian Life Course Survey (HLCS) longitudinal dataset drawn from the population of 8th graders students who completed the Hungarian National Assessment of Basic Competences (NABC) in May 2006. The HLCS original sample was 10,022 students with valid NABC test scores and students with special education needs who completed a simplified version of the reading test. The students were surveyed yearly from 2006 to 2012 with the exception of 2010. The last survey wave has a sample of around 7,000 students. Even though attrition occurred at each survey wave, it does not seem to be systematic (see Simonovits and Kezdi, 2016). The first two survey waves were also used in Kertesi and Kezdi (2011, 2014) while Simonovits and Kezdi (2016) use all six waves. This thesis uses all six waves and benefits greatly from the richness of the datasets. The data provides extensive information on family background, health, socio-emotional skills, attitudes, school outcomes and others. Students with low test scores were oversampled in order to gather a large enough sample of Roma students. Throughout my analysis I use sampling weights for national representativeness.

The sample used throughout the analysis is 5,223 observations due to attrition and lower response rates for some of the dependent variables. Attrition is not systematic as shown by the summary statistics for the main sample and full sample presented in Tables A10-A12 in the Appendix⁴.

4.2. Dependent Variables

Socio-emotional skills

The dependent variables used cover various dimensions of socio-emotional skills, at various ages, as the child approaches adulthood: self-esteem, locus of control, social

⁴ The students in the main sample have higher skills and better socio-economic status but the difference is small

competences, expectations, life satisfaction and depression. There is no consensus in the literature regarding the optimal measures of socio-emotional skills. The most widely used classification of personality traits is the Big Five: conscientiousness, openness to experience, extraversion, agreeableness and neuroticism (emotional stability). Each of the five domains has several subdomains. However, the Big Five has received criticism and several alternatives have been proposed (see Almlunnd et al., 2011). The lack of consensus regarding optimal measurement is reflected also in the fact that most of the papers cited in my literature review use different measures of socio-emotional skills.

The observed variables used to construct the main variable are self-esteem, locus of control, social competences and depression at age fifteen. These variables are part of two domains of the Big Five: neuroticism (emotional stability) and extraversion, thus reflecting both internalizing and externalizing behavior. I estimate the latent socio-emotional skills variable using maximum likelihood factor analysis estimation⁵. I use only one time observation for each measure because otherwise the estimation generates Heywood cases. Expectations and life satisfaction are not included because they do not necessarily reflect socio-emotional skills⁶. Using more measures to construct the main dependent variable has the potential to reduce measurement error (Attanasio, 2015). The factor analysis retains only one factor with eigenvalue greater than one which generates the main dependent variable. Summary statistics of the indexes and their items are available in Tables A1-A7 in the Appendix.

Self-esteem is measured using a modified version of the *Rosenberg's Self-Esteem Scale* (Rosenberg, 1965) during the first and forth waves. The scale has ten Likert items reflecting the subjective appreciation of oneself as a person. It is a widely used measure in longitudinal studies

⁵ The factor loadings are presented in Table A9 in the Appendix

⁶ Expectations regarding the future might reflect the optimism of the individual but might just as well be a reflection on ones skills or perspectives for the future.

and was also used in Heckman et al. (2006) and Cunha and Heckman (2010). Gertler et al. (2014) use self-esteem to construct a measure of internalizing behavior.

The locus of control is measured using an abbreviated four item *Rotter's Locus of Control Scale* (Rotter, 1966) during the first and second waves. The locus of control reflects to what degree a person believes that events taking place in his/her life are the result of their own actions (internal locus) or not (external locus). Rotter's Locus of Control is widely used as well and is part of the socio-emotional skills measurement in Attanasio et al. (2015) and Heckman et al. (2006).

Social competences are measured during the first and third waves and use the Harter's Social Competence Scale (Harter, 1982). The scale reflects child's peer relations and one's own perception in the eyes of his/her peers. Most of the papers cited above use similar measures of sociability.

I construct a measure of depression using a five item scale from the first survey wave. Depression is a subdomain of the neuroticism component of the Big Five. Attanasio et al. (2015) uses depression as a measure of mother socio-emotional skills.

Life satisfaction is measured in the first and sixth survey, children assessing how satisfied they are with their lives, on a scale from one to ten in the first wave (age 15), and answering a yes or no question on whether they are satisfied with their lives in the sixth wave (age 21).

Expectations are constructed using series of questions during the first and sixth wave, when children are asked to state the subjective probability of certain positive or negative events taking place in their life in the near or far future. The stated probabilities are categorized as very low (less than 25%), low (25-50%), high (50-75%) or very high (more than 75%) expectations and then aggregated.

All the variables are standardized using the weighted means and standard deviations to ease the interpretation of the regression coefficients. If less than 30% of the answers to the items

13

of an index are missing, I replace the missing answer with the sample mean for the item⁷. Shrive et al. (2006) and Downey and King (1998) show that this procedure is reliable for less than 30% missing items.

Cognitive skills and Health

As proxies for cognitive skills, I use the NABC test scores, averages in mathematics, Hungarian literature and language, foreign language and the 8th grade's average. I acknowledge the fact that school averages might reflect more a mixture of crystallized intelligence and socioemotional skills (motivation, self-confidence, perseverance) (see Almlund et al., 2011). The NABC test scores were intended to measure how well children could use their skills in real life situations. In this sense it could be argued that they could reflect fluid intelligence more than school results even though also imperfectly. Unfortunately there is no widely accepted measure of cognitive skills in the data. I report the results for all measures. I estimate the latent variable reflecting academic achievement through factor analysis and use as measurements all the school averages and the reading NABC test score. The mathematics NABC test score is dropped from the factor model because it generates a Heywood case.

Health is measured through a self-reported evaluation at age 15. Children evaluate their health as either poor, fair, good or excellent.

4.3. Explanatory Variables

Early Childhood Investment

The independent variable of interest is early childhood parental investment which I mostly refer to as early childhood investment or early childhood parenting. The variable is constructed through maximum likelihood factor analysis, using answers reported by the children and parents, during the first wave of the survey, to various retrospective questions indicating the frequency of various child stimulating interactions (e.g. bedtime stories, concerts)

⁷ The results are similar if using individual average on valid items

when the child was at preschool age. The survey questions are presented in Table A7 in the Appendix

Having both parents and children answer these questions could reduce potential biases (social desirability or recall bias). When performing the factor analysis, I drop the variables reflecting the frequency of radio or CD story listening and TV story watching because they have very low factor loadings.

The variables reflects parental time investment. However, in the absence of data on early childhood material investment, the coefficient could reflect the effect of both time of investments. It is likely that parents who spend time doing various activities with their children also provide them with other inputs. Attanasio et al. (2015) is the first paper, to my knowledge, to differentiate between parental time investment and material investment based on factor loadings on investment variables. Using evidence from a randomized control trial they find that time investments impact more the accumulation of socio-emotional skills, while material investments matter more for cognitive skills. In one specification, I try to provide some insights on this by adding controls for contemporaneous material investments (books, internet etc.) as proxies for early childhood material investment. While imperfect, I think it is plausible to assume that these proxies provide some credibility and insights.

Environmental factors: socio-economic status

There is vast amount of evidence on the strong relation between socio-economic status and human capital development (McLoyd, 1998). Poor children receive significantly less homebased cognitive stimulating experiences, warmth and support, and usually face more authoritarian parents and domestic violence (Evans, 2004). Rubio-Codina et al. (2014) indicate that socio-economic status is associated with cognitive and language delays that develop as early as the time the child is 12 months old and grow significantly in time. Furthermore, socioeconomic status is a strong predictor of child health, affecting the child even before birth through the mother's nutritional status with long term consequences on child development (Barker, 1995). After birth, nutrition also seems to be very important. Kramer et al. (2001) suggest that breastfeeding is associated with subsequent positive outcomes while Victora et al. (2008) summarize the substantial positive results of long term early childhood nutrition interventions.

There are competing theories regarding the impact of poverty on parenting practices. Less educated parents might have wrong beliefs on what optimal parenting represents (Attanasio, 2015) suggesting that information might be particularly important for improving parental stimulation. On the other hand, recent evidence from behavioral economics, points out that poverty can affect the mental capacity of individuals to elicit attention to other issues except immediate pressing concerns (Shah et al., 2012; Mani et al., 2013). Thus, poor parents could be less capable of focusing their attention on children. In this framework, poor people make worse parents not necessarily because they are unaware or have wrong beliefs about what represents good parenting practices, but because poverty affects their capacity of being attentive to children's needs when other concerns appear to be more urgent. The two theories suggest completely different directions of intervention. In this thesis, I am unable to test which theory fits the data, but both suggest that socio-economic status is associated with poorer parenting practices.

The data provides extensive information on family background, such as the education, employment status, income and health of the parents. Whether the child has Roma ethnicity is also controlled for. Similar to Kertesi and Kezdi (2011 and 2015), I consider a child to be Roma if one of the biological parents declares its first or second nationality to be Roma in the first and second waves of the survey. In addition to this, I consider a child Roma also if she/he declares herself/himself to be Roma in the fourth or sixth waves of the survey. Besides income, I include other measures reflecting the poverty status of the family, like whether it happened in the past year to not have enough money for food, rent or heating, whether the child receives social assistance at school (free or discounted meals, textbooks, education aid), whether the parents receive social benefits (unemployment, childcare, pension) and whether the household has certain utilities (bathroom in the house, computer, car).

Furthermore, household structure and information are also included, controlling for the number of individuals in the household, apartment size per person, number of rooms per person and the number of individuals in the household with various occupational statuses. I also control for the total number of siblings which could have a negative effect on parental investment because parents might have to shift their attention between siblings or because they could pass this responsibility to older children. On the other hand, having more siblings could result in a more stimulating environment at home, thus improving skills. Also, if parents have more children, the mother or father benefiting from childcare might spend more time at home with the children.

I include also fixed effects for the quality of the neighborhood, region of residence in Hungary, and whether the family lives in Budapest or other large cities, other smaller cities or in a village.

Early childhood health

Early childhood nutrition and health predicts a great variety of outcomes later in life (Victora et. al, 2008). In this analysis, I am trying to detect the dynamics of parental investment choices in response to the health of the child in the early years. Do parents compensate or reinforce early health statuses? For this, I control first for a rich set of variables reflecting early childhood health: weight at birth, standardized height for age as a proxy for stunting, and whether the child suffered from various health issues.

Finally, the model also controls for the gender of the child to observe if there is any evidence of parental investment preferences in early life for girls or for boys.

School and class fixed effects and home environment

The next set of variables are included in the regression to try to separate the effect on skills formation of early childhood investment from investments at different periods in the life of the child, or from other forms of investment (e.g. material investment). This might generate several problems because these variable raise similar identification issues as the early childhood investment variable. These issues will be further discussed in the next section.

School and teacher quality is expected to have a direct relationship with skills development, while also reflecting parental investment. Parents who try to provide their children with better educational environments at school, potentially might have also invested more time with their children in early years. Or parents might respond to children's skills by trying to offer them better education quality to try to reinforce or compensate those skills. In the main sample I am unable to use class fixed effects because the number of students per class is small. However, I use class fixed effect in the robustness checks for the full sample available in column (7) of Table A13 in the Appendix. The coefficient on early childhood investment stays fairly constant.

The model controls for the level of parental investment at age 15 through the HOME index used in Kertesi and Kezdi (2011, 2015). The HOME index is a widely used measure of home environment quality and parenting, with significant predictive power of various school outcomes. The scale used was adapted for teenagers and includes a cognitive subscale and emotional subscale, which I introduce separately in the regression to identify their association with socio-emotional skills, cognitive skills and health.

Furthermore, I try to proxy for parental beliefs with respect to the production function using a variable describing the level of education they aspire their child to reach. If parents do not know or have potentially wrong beliefs about the return to education, they might choose to invest less in their children. In the same time this variable could describe the response of parents to the development status of their children. If parents are trying to maximize some objective function they might choose to invest more resources in children which show more potential.

I also control for other variables potentially reflecting parenting practices, such as whether the child used to work at primary school age or was working during the first survey wage (age fourteen) for the family farm or business, whether the child reports during the sixth survey wave being abused (psychologically, physically or sexually) by relatives, friends or others before they were 14 or after.

Finally, the model controls for the presence of stimulating items in the household: number of books, internet connection at home, whether the child owns at least one book (not textbooks), whether she/he has her/his own room, desk or computer. These variables can be viewed as proxies for material parental investment.

The description and coding of all the explanatory variables are available in Table A8 of the Appendix. The model also includes dummy variables for missing observations in each independent variable.

Potentially omitted variables

Even though the model controls for numerous variables, omitted variables bias is very likely to be present. First of all, besides early childhood health, there is no information in the data regarding early childhood socio-emotional skills or non-cognitive skills. Parental investment decisions might be endogenous to these skills. This would not be an issue if early childhood health, cognitive and socio-emotional skills would be perfect substitutes but this is a very strong assumption. However, these skills are strongly directly correlated during adolescence suggesting that at least partially the issue might be attenuated. Very likely, there are other unobserved early childhood development shocks that impact parental investment decisions that we do not observe. This issue is discussed in the following section.

5. Empirical Model

The benchmark approach for estimating the human capital function proposed by Cunha et al. (2010) and extended by Attanasio et al. (2015), uses structural models to account for the dynamic complementariness between the factors of the human capital and its various inputs. Unfortunately in this analysis, I estimate a reduced form model being constrained by the availability of the data. The data has information on early childhood parental investment and health at only one time period, and includes measures cognitive and social skills only when the children are already teenagers or approaching adulthood. This makes the identification difficult and I cannot infer causality from the analysis. However, I control for a wide range of factors which I introduce sequentially to highlight the changes in size and significance of the coefficient on early childhood investment.

In the first specification I include only the investment variable. I then control for environmental factors. Following this, I try to provide some insights on the parental investment decision process by introducing early childhood health in the model. If parents compensate for early shocks and differences, the coefficient on investment is expected to increase, while if their behavior is reinforcing the coefficient is expected to decrease in size. I also look at whether parents display a preference for investing more in boys or girls. Finally as an extension to the model, I control for other endogenous environmental factors reflecting the home environment, school quality, abuse, child labor and others. While these variables raise similar endogeneity issues as the early childhood investment variable, it also separates the effect of early investment from contemporaneous investment and other forms of investments.

I estimate OLS regression models where I allow the dependent variable (the dimensions of human capital $\theta_{i,T2}^{k}$ - socio-emotional $\theta_{i,T2}^{S}$, cognitive $\theta_{i,T2}^{C}$ and health $\theta_{i,T2}^{H}$) during youth to depend on early childhood parental time investment (*X_{i,t}*), contemporaneous socio-economic

status ($Z_{i,T2}$), early childhood health ($\theta_{i,T1}^H$) and gender (*G*). The main specification has the following shape:

$$\theta_{i,T2}^{k} = \beta X_{i,t} + \gamma_1 Z_{i,t2} + \gamma_2 \theta_{i,T1}^{H} + \gamma_3 G + e_{i,T2}$$
(8)

The error term could be further decomposed into child endowment or heterogeneity $(\varepsilon_{i,T2}^{p})$, unobserved family heterogeneity and preferences $(\varepsilon_{i,T2}^{F})$ and idiosyncratic shocks $(\varepsilon_{i,T2})$:

$$e_{i,T2} = \varepsilon_{i,T2}^p + \varepsilon_{i,T2}^F + \varepsilon_{i,T2} \tag{9}$$

Several identification issues of the β coefficient arise from the endogeneity of the parental investment. Firstly, parental investment might be endogenous to family preferences. Parents who provided higher level of early childhood investment, might also invest more, later in life and also provide their children higher levels of other inputs such as better nutrition, stimulating play materials and high quality education. In this situation the coefficient will be upwardly biased. I try to attenuate this issue by estimating also the specification in Equation (10), where I control also for a reach set of variables reflecting the home environment and parental preferences ($M_{i,T2}$), and by including school fixed effects. However, this raises certain issues since the added variables are chosen by parents thus likely endogenous in a similar way to the early childhood investment variable.

$$\theta_{i,T2}^{k} = \beta X_{i,t} + \gamma_1 Z_{i,t2} + \gamma_2 \theta_{i,T1}^{H} + \gamma_3 M_{i,T2} + \delta SchoolFE + e_{i,T2}$$
(10)

The second identification issue is even more problematic. This stems from the reinforcing/compensating behavioral mechanisms discussed earlier. Parental investment decisions might be endogenous to the child endowment error term. Child fixed effects would not eliminate this concern because it is unlikely that the expression of child endowment is

constant over time. I try to deal with this issue by including a reach set of controls reflecting early childhood health. There is no information regarding early childhood cognitive and socioemotional skills in the data.

However, in spite of the controls I acknowledge that there might be other factors not captured in the model correlated with both investment and skills. Attanasio et al. (2015) emphasize that accounting for the endogeneity of the investment function is crucial. In their analysis not accounting for endogeneity generates a severe, fivefold downward bias of the coefficient. Helmers and Patnam (2011) results point in the same direction even though the bias is smaller in size. Both papers use instrumental variable estimations. Attansio et al. (2015) use average female and male wages in the village, household wealth and mother's marital status as instruments and argue that the exclusion restriction is likely to hold since the model also controls for numerous child and parents characteristics. They note that according to economic theory, exogenous changes in household resources should impact investment only through the budget constraint and does not enter the production function directly. Helmers and Patnam (2011) use birth order as instrument for investment.

Benefiting from a rich database, I also try to implement an instrumental variable strategy. I identify three sets of variables as potential IVs even though the exclusion restrictions pose some issues. One first candidate reflects birth order through the number of younger children in the household when the child was 6 years old and the number of children in the household less than 3 years older than the child⁸. More children in the household might cause material and time resources allocation (or reallocation) across children. A kindergarten age child might receive less attention and play materials if the parents have infants to attend. On the other hand, more siblings can create richer social interaction which would have a direct impact on cognitive

⁸ I avoid using both variables in the same model because of collinearity concerns with the total number of siblings variable. The results are very similar if either variable is used especially in the full sample.

or socio-emotional skills. Moreover, families with higher number of children might be poorer and have less educated parents. I hope these concerns are attenuated through the variables on total number of siblings, household structure, and parents' wealth and education included in the main model.

The second IV candidate is whether the child did not live or was separated from one of the parents before school age and whether a stepparent started living with them. Single parents might have to work harder to provide for the family and might have less available time to spend with the children⁹. This acts as a shock to household resources with an effect on child development through parental investment. It can be argued that this is correlated with the family heterogeneity error term but again I hope that all the family background control deal with this issue.

The third set of IV variables reflect employment history of the parents during the time the child was at preschool age. I create dummy variables for whether the parents were employed during the time child was 3-6 years old¹⁰. Employment variation causes fluctuations in household resources and material and time investment in the child. Exclusion restriction would not hold if parents would change their labor supply as a result of a child development shock. For example, having a sick child might force the parents to work more to provide the resources the child needs. While I acknowledge this to be possible, I am hopeful that the rich set of controls reflecting early health, household health and structure, and parents characteristics to mitigate this issue. I avoid using wealth variables as IVs as Attanasio et al. (2015) as I find exclusion restriction to be highly unlikely.

⁹ The main reasons for separations were either divorce or death of one of the parents, or in the case of separation from father also the fact that the father never lived with the family

¹⁰ Using the number of working years instead of the dummy generates similar results

6. Regression Results

6.1. OLS regressions

Early childhood investment coefficients

I turn now to the regression results and begin with the coefficients of the main independent variable, early childhood investment, illustrated in Table 1, on socio-emotional skills, cognitive skills and health. As indicated in the previous section, I use a methodology similar to Kertesi and Kezdi (2011, 2014) and include groups of variables sequentially to provide potential insights on the causal mechanisms. I start with a single variable model on early childhood investment, followed by a rich set of variables describing the socio-economic status of the household, early childhood health, gender, home environment and school fixed effects. I am interested on how the coefficient of interest changes in size and significance with the inclusion of the sets of controls.

The order of the controls is motivated by the identification strategy. I first control for the socio-economic status of the household because I expect poorer families to provide less stimulation to the children. Then, the model controls for early childhood health which, conditional on family background, is expected to be exogenous. At this stage I try to identify if there is any sign of compensating or reinforcing behavior based on the change in the coefficient of interest. In the next step, gender is included to investigate any gender biased investment preferences. As emphasized in the previous sections, the last two specifications are more problematic because they are likely endogenous to the child heterogeneity error term. School quality and other variables reflecting contemporaneous and past home environment might be chosen by parents potentially as a response to child endowment. But, including them in the model provides some insights into the relative importance of different types of investment (material or in time) and timing of investment on skills development.

	(1)	(2)	(3)	(4)	(5)	(6)
		Socio-	Early	Gender	Home	School Fixed
		Economic	Childhood		Environment	Effects
		Status	Health			
Socio-	0.197***	0.152***	0.155***	0.158***	0.123***	0.121***
emotional	(0.014)	(0.014)	(0.014)	(0.014)	(0.015)	(0.015)
	0 108***	0.087***	0.080***	0 088***	0.060***	0.060***
Satisfaction	(0.020)	(0.037)	(0.014)	(0.014)	(0.015)	(0.015)
Age 15	(0.020)	(0.014)	(0.014)	(0.014)	(0.015)	(0.015)
Life	0.038***	0.008	0.009	0.008	0.000	0.001
Satisfaction	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Age 21					× ,	. ,
Self Esteem	0.160***	0.140***	0.143***	0.147***	0.121***	0.115***
Age 15	(0.017)	(0.016)	(0.016)	(0.016)	(0.016)	(0.017)
Self Esteem	0.088***	0.066***	0.069***	0.073***	0.055**	0.047**
Age 18	(0.021)	(0.022)	(0.022)	(0.022)	(0.022)	(0.023)
External Locus	-0.103***	-0.062***	-0.066***	-0.065***	-0.034*	-0.037**
of Control Age						
15	(0.015)	(0.015)	(0.015)	(0.015)	(0.017)	(0.018)
External Locus	-0.090***	-0.062***	-0.062***	-0.063***	-0.045**	-0.042*
of Control Age						
16	(0.020)	(0.021)	(0.022)	(0.022)	(0.022)	(0.022)
Social	0.137***	0.114***	0.115***	0.115***	0.095***	0.095***
Competences						
Age 15	(0.016)	(0.017)	(0.017)	(0.017)	(0.016)	(0.016)
Social	0.092***	0.088***	0.087***	0.088***	0.079***	0.080***
Competences						
Age 17	(0.017)	(0.018)	(0.018)	(0.018)	(0.017)	(0.018)
Expectations	0.294***	0.153***	0.153***	0.152***	0.098***	0.090***
Age 17	(0.024)	(0.019)	(0.020)	(0.020)	(0.018)	(0.020)
Expectations	0.273***	0.105***	0.102***	0.102***	0.059***	0.052***
Age 21	(0.021)	(0.017)	(0.017)	(0.017)	(0.016)	(0.016)
Depression	0.190***	0.123***	0.124***	0.117***	0.068***	0.059***
Åge 15	(0.019)	(0.020)	(0.019)	(0.019)	(0.019)	(0.019)
Learning Age	-0.149***	-0.103***	-0.104***	-0.107***	-0.079***	-0.080***
16	(0.013)	(0.014)	(0.014)	(0.014)	(0.015)	(0.016)
Academic	0.241***	0.090***	0.090***	0.080***	0.012	0.010
Achievement	(0.019)	(0.016)	(0.015)	(0.014)	(0.013)	(0.014)
Age 15						
Math Score	0.237***	0.053***	0.051***	0.054***	0.004	-0.001
Age 15	(0.022)	(0.017)	(0.017)	(0.017)	(0.016)	(0.016)
Reading Score	0.240***	0.066***	0.064***	0.057***	0.005	0.003
Age 15	(0.020)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Health Age 15	0.053***	0.025**	0.027***	0.028***	0.019*	0.019*
- excellent	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)

Table 1. The coefficient on early childhood investment as groups of controls are added using the variables on the rows as dependent variables

*** p<0.01, ** p<0.05, * p<0.1.Clustered standard errors at school level in brackets. Weighted by sampling weights. The sample is 5231 observation in all models except the math score model with 4991 observations due to missing score for special education needs children.

First of all, we notice that the coefficient on early childhood investment remains highly significant after including the full set of controls in all of the models except the life satisfaction at age 21¹¹ and cognitive skills models. I discuss the changes in the coefficient and other coefficients in the model in the following subsections. The full set of regression coefficients for columns (4) and (6) are available in Tables A14-A15 in the Appendix.

In the socio-emotional skills models, after the inclusion of controls for socio-economic status, home environment and school fixed effects, the coefficient of interest drops in size but still indicates a moderate association between early childhood investment and socio-emotional skills. On the other hand, when introducing early childhood health in the model, the coefficient increases in absolute terms, suggesting a negative correlation between early childhood health issues and parenting. As we shall see from the first stage regressions in the IV models, children with various health issues during early childhood had significantly higher levels of parental investment. This could be interpreted as evidence of compensating behavior with respect to early childhood endowment. Controlling for gender also increases the size of the coefficient. Girls received slightly higher parental investment but I cannot infer much about parental gender preferences.

The coefficient in the main model with the dependent variable - the estimated socioemotional latent variable- is 0.12, indicating that one standard deviation increase in early childhood investment is associated with 0.12 higher levels of socio-emotional skills after controlling for socio-economic status, early health, gender, home environment and school. In the models with observable dependent variables the coefficients range between 0.05-0.1 in absolute value indicating a slightly weaker association but still significant. We also observe that the effect seems to slightly fade out in time as the children grow up.

¹¹. Life satisfaction at age 21 is measured as a dummy variables and has much less variation than life satisfaction at age 15 which is measured on a scale from one to ten.

In the cognitive skills regressions the coefficient drops even more after controlling for socio-economic status and becomes insignificant after we control for the home environment. Other forms of investment and timing of investment seem to matter more for cognitive skills. In the health regression the coefficient is small even in the single variable model and drops slightly when controls are added. Even though the coefficient remains significant at 10% significance level, the effect is very small. One standard deviation increase in investment is associated with 2 percentage points higher probability that the child reports having excellent health during adolescence. The model does not have high explanatory power, suggesting that there are factors impacting health that are not part of specification such as nutrition or access to good quality medical services.

Transmission mechanisms

Following a methodology similar to Kertesi and Kezdi (2015), I try to identify the transmission mechanisms behind the changes in the coefficient on early childhood investment. In Table 2, I present the changes in the estimated coefficient when including sequentially the groups of controls first (upper bound) or last (lower bound). The questions I try to answer through this exercise are: is the decrease in the size of the coefficient of interest caused primarily by socio-economic status, or by the fact that parents who invested time in their children in early childhood also provided other inputs? Do high early investment parents also try to provide better schooling? How does the coefficient change after controlling for early health or gender; is there any evidence of parental preferences?

Dependent Variable		Socio- economic status	Early Health	Gender	Home Environment	School Fixed Effects
Socio-emotional	lower bound	0.006	-0.004	0.001	0.037	0.002
skills	upper bound	0.045	-0.002	-0.001	0.074	0.000
Academic Achievement	lower bound	0.005	0.000	-0.001	0.070	0.002
	upper bound	0.151	0.006	0.003	0.228	0.012
Health	lower bound	0.004	-0.002	0.000	0.010	0.000
	upper bound	0.028	-0.002	0.036	0.033	0.024

Table 2. Changes in the coefficient on early childhood investment as groups of controls are added first (upper bound) or last (last) in the regression model

The results from Table 2 indicate that after controlling for home environment, early childhood health, school quality and gender, no significant difference in early childhood parenting is observed between families with different socio-economic backgrounds. However, the quality of the home environment appears to be strongly associated with early childhood parenting even after controlling for socio-economic status. Parents who provide more early childhood stimulation to their children are also more likely to provide other forms of investment and at other times in their life, regardless of socio-economic background. Controlling for school fixed effects does not change the coefficient on early parenting by much.

Early childhood health appears to be negatively associated with early childhood investment suggesting that parents might try to compensate through higher investment poorer health early in life. With respect to gender, there is no clear evidence of gender biased parental preferences.

Coefficients on control variables

In the remaining part of this subsection, I discuss briefly the results on the other variables in the full specification of the models on socio-emotional skills, cognitive skills and health, with some references from the other models. I begin with the coefficients from the socio-emotional skills models, followed by cognitive skills and health. Roma students report 0.08 standard deviations higher levels of socio-emotional skills but the results are significant at 10% significance level only in the full sample. However, Roma students are more satisfied with life at age 15, but significantly less satisfied at age 21 and have lower expectations about the future. While we can only speculate, the former result could be explained by potential higher social capital amongst Roma communities, while the latter could reflect the fact that Roma children are aware they are discriminated against thus having lower hopes for the future. This is supported by the fact that the coefficient in the expectations model more than doubles in size to about 0.25 standard deviations between age 17 and 21 suggesting that, as Roma children grow up, they might face ethnicity related problems in society and become less optimistic about the future. The difference in academic achievement is small and only significant at 10% in the main sample but, as reported in Kertesi and Kezdi (2011, 2015), Roma students have lower scores in mathematics. With respect to health, Roma students are 8 percentage points more likely to report having excellent health even after controlling for all the other factors.

A large share of the variables describing the socio-economic status of the family are statistically insignificant in the socio-emotional skills model¹². They are stronger predictors of life satisfaction, expectations and learning style. While insignificant in the socio-emotional skills regression, the education level of the parents (potentially reflecting the quality of child-parent interaction and genetic inheritance) is a strong predictor of academic achievement and test scores. Various family income and household structure variables also seem be important predictors of these variables. Socio-economic status is a weak predictor of current reported health. Actually, the only strong predictors of current health in the model are early childhood health and history of abuse.

¹² The association is stronger with respect to poverty indicators and the quality of the neighborhood

Early childhood health is strongly associated with socio-emotional skills and health during adolescence. Children who suffered from various illnesses during early childhood have significant lower development statuses¹³. Low weight at birth strongly predicts the proxies for cognitive skills during adolescence but not socio-emotional skills or health. This result is consistent with the remarks from Attanasio (2015) and Heckman et al. (2013) which point out that long lasting effects on cognitive skills are mostly achieved through interventions in the first 3 years of life. Birth weight reflects early childhood health but could also proxy for the quality of prenatal care and parenting during the first years of life.

The home environment proves to be particularly important for the development of socioemotional and cognitive skills. With respect to socio-emotional skills, what matters most is the quality of parent-child interactions and less the presence of stimulating objects in the household. The two HOME indexes account together for an association similar in magnitude to early childhood parenting. Henceforth, socio-emotional skills could potentially be improved through interventions even later in life¹⁴. In the cognitive skills models, the association with the HOME indexes is less clear. Only the HOME cognitive index seems to matter for the estimated cognitive skills variable and test scores, while the HOME emotional index has a negative coefficient on test scores¹⁵. In the health model both coefficients are insignificant.

What seems to be very important for the cognitive skills proxies but not for emotional skills is the presence of stimulating objects in the household (number of books, whether the child owns any non-textbook books, internet at home etc.). These could be viewed as a form of material investment. Even though there is no information on whether they were available also

¹³ In particular, speech and hearing impediments, respiratory problems and whether the child was abnormally thin or weak are strongly associated with lower socio-emotional skills later in life. ¹⁴ In the other models on observable socio-emotional skills, there is evidence that the HOME emotional index matters more for various measures of socio-emotional skills than the HOME cognitive index.

¹⁵ One potential explanation for this could be that parents try to compensate thus providing more emotional support to children with poorer skills.

in early childhood, one could expect a family with stimulating objects in the household to have had provided a similar environment during the child's early childhood. This result is consistent with Attanasio et al. (2015) which find that time investment are more important for socioemotional skills while material investment matter more for cognitive skills.

Parental aspirations regarding the desired completed level of education of the child, our proxy for parental beliefs regarding the production function, strongly predicts the development of socio-emotional and cognitive skills. Children whose parents aspire that they graduate high school or professional training/higher education, display significantly higher levels of skills than children whose parents only aspire they will finish at most 8 grades¹⁶. Parents with higher educational aspirations also display larger levels of early childhood investment. I cannot establish whether this is a sign of lack of information, or parental altruism, or a response of parents to their children development, but it raises an important research question.

In addition, whether the child reported being abused psychologically, physically or sexually by relatives, friends or strangers accounts for around 0.3 standard deviations of the differences in socio-emotional skills¹⁷. Being subjected to severe emotional stress affects various levels of socio-emotional skills such as self-esteem, locus of control, social competences, expectations and correlates with lower levels of life satisfaction and higher levels of depression. Finally, my measure of child labor, whether the child used to work on the family farm or business when he was of elementary school age, is associated with lower socio-emotional and cognitive skills during adolescence.

As mentioned before, there are identification problems in the analysis and I cannot make causal claims. First of all, it was not possible to estimate a panel model which would have potentially eliminated some of the heterogeneity between children and would have allowed

¹⁶ 0.2/0.4 sd in the socio-emotional skills model, and 0.3-1 sd in the cognitive skills models

¹⁷ No association with cognitive skills

more variations in parenting at different points in time. As emphasized in the literature, human capital development is a very dynamic process, past level of human capital making investments more or less productive in the future. While there is only one time observation of early childhood parenting, the HOME indexes could be a reflection of the long term parenting quality and home environment. What the analysis tried to identify primarily was: the difference in skills, from adolescence and early adulthood, between children with similar socio-economic background, home environment, early childhood health and from the same schools, but which were exposed to different levels of stimulating activities by the parents in early childhood.

Even after controlling for all the variables in the model, early childhood investment could still be endogenous to various factors. There is no measure of early childhood socio-emotional skills or cognitive skills in the data. Parents could choose to compensate or reinforce with respect to these skills as well. This would not be a serious concern if socio-emotional and cognitive skills would substitute with health. While this assumption is strong, one could expect children with health problems in early years to have less opportunities to develop emotional or cognitive skills as well. The following section tries to deal with these issues through instrumental variable estimation models.

6.2. Instrumental variable regressions

In this final section, I discuss the estimation results from the instrumental variables regression models. I begin with a brief discussion of the first stage coefficients followed by the second stage results. The instruments I use are: the number of younger children in the household when the child was 6, whether the child stopped living with one of the parents before he was 7, and whether a stepparent started living with them during this period. I drop the employment status of the parents in early childhood as a potential IV because it is insignificant in the first stage regression and lowers the joint significance of the instruments to a significance level of

10%. As emphasized earlier, the empirical strategy could raise some potential concerns and does not fully solve the identification problem. Even though the instruments are jointly significant they are still relatively weak instruments. While some caution is needed when interpreting the coefficients, the instrumental variable estimation suggests a large downward bias in the OLS models and gives some insight on parental behavior.

The IV estimation results illustrated in Table 3 use the specification in Equation (8) which controls for socio-economic background, gender and early childhood health. I avoid controlling for the home environment variables and school fixed effects because of endogeneity concerns. These variables, like early childhood investment, are chosen by parents making them endogenous to similar factors. Including them in the IV regression would likely reintroduce the bias in the first stage regression.

I briefly present the first stage results before proceeding to the main results. In Table 3 only the coefficients on the instruments are presented. The full regression results are available in Table A *17* in the Appendix. Firstly, I turn to the coefficients on the instruments. Having younger siblings in early childhood is positively associated with higher levels of investment. This could be explained by potentially richer social interactions in the household, shared resources (toys, books etc.) and by the fact that at least one parent probably spent more time at home. Being separated or not living together with the biological mothers has a large and negative association with investment. In the full sample, not living with the father also has a negative and significant coefficient but smaller in size¹⁸. Also, in the full sample, a father living with the family or joining the household in early childhood is strongly associated with higher levels of investment, potentially due to higher emotional and financial resources in the family.

The other variables from the first stage regression which are associated with significantly higher levels of investment are: health issues during childhood, mother education level, lack of

¹⁸ A much larger share of children did not leave or stopped living with the father in early childhood

poverty and neighborhood quality. Children who suffered from various health issues during early childhood received higher stimulation from their parents which indicates compensatory behavior. The education level of the mother strongly predicts early childhood investment¹⁹. Cunha et al. (2010) find that parents' socio-emotional skills are strongly associated with investment, while Attanasio et al. (2015) find that mother's cognitive skills has a strong effect on child's skills but mainly through investment. While there is no measure of mother's skills in the data, my results could be interpreted in a similar fashion. Mother's education level is associated with socio-emotional skills only through investment, but correlates with the cognitive skills proxies also directly. Finally, there is no difference in terms of investment between Roma and non-Roma children in the main sample after controlling for socio-economic status and early childhood health.

	OLS	IV	First Stage Regression – Early Investment	Childhood
Socio-Emotional	0.158***	0.510**	Younger siblings at age 6 ^a	0.046*
Skills	(-0.014)	(0.251)		(0.024)
Academic	0.080***	0.487**	Separated from mother Age 0- 7	-0.334**
Achievement	(-0.014)	(0.233)		(0.129)
Math Saara	0.054***	0.388	Separated from father Age 0-7 ^b	-0.077
Mail Scole	(-0.017)	(0.271)		(0.059)
Uaalth avaallant	0.028***	0.000	Stepfather Age 0-7 ^b	0.198
Health - excellent	(-0.01)	(0.124)		(0.125)
			Stepmother Age 0-7	-0.051
				(0.204)

Table 3. IV	Regression	Results
-------------	------------	---------

*** p<0.01, ** p<0.05, * p<0.1.Clustered standard errors at school level in brackets. Weighted by sampling weights. The sample is 5,231 observation in all models except the math score model with 4,991 observations due to missing score for special education needs children. Controls for socio-economic status, early childhood health and gender included

^a Insignificant in the full sample estimation.

^b Statistically significant in the full sample estimation

¹⁹ Children who's mother's completed 0-8 grades/vocational/secondary school have 0.4/0.2/0.1 sd lower levels of investment

Similar to Attanasio et al. (2015) and Helmers and Patman (2011) the coefficient on early childhood investment increases almost three fold to 0.51 standard deviations in the socioemotional skills IV estimation and six fold to 0.49 standard deviations in the academic achievement IV estimation. But it is worth noting that, as in Attanasio et al. (2015), the precession of the estimate is much smaller because the instruments are not very strong. In the mathematics test score model, the coefficient is larger but not statistically significant.

As emphasized before, the IV strategy does not fully solve the identification issue. On one hand, I assumed that the instruments, conditional on socio-economic status and early childhood health, affect the development of the child only through investment. However, the exclusion restrictions might not hold under certain conditions discussed in the final part of section 5. On the other hand, even though I can reject under-identification at high significance level, the instruments do not explain a large share of the variation in investment. The hypothesis would not hold if I would have used only one group of instruments. In effect, the standard errors are larger and the 95% confidence intervals for the estimates are wide.

Even though I do not place high confidence on the point estimate, the direction and size of the potential downward bias in the OLS model provide important insights. Early childhood parental investment plays a very important role for the development of the child and it appears to have a persistent effect, especially on socio-emotional skills. Not taking into account the endogeneity of investment generates a large downward bias suggesting that parents try compensate if the development of the child is affected by negative shocks.

7. Conclusion

This thesis investigated the impact of early childhood parental investment on socioemotional skills, cognitive skills and health, later in life, using a longitudinal dataset following 10,000 students for 6 years from age 15 to 21. The dataset provides also retrospective information regarding early childhood parenting practices, health and various family background characteristics.

I estimated OLS models for socio-emotional skills, cognitive skills and health (estimated latent and observed variables) on early childhood parental investment and controlled for a rich set of controls for socio-economic status and early childhood health. One limitation of the study is the absence of information regarding early childhood socio-emotional and cognitive skills to be used as controls in the analysis.

Acknowledging this limitation and the likely endogeneity of investment to unobservable child endowment's shocks, I estimated also IV models where the key variable is instrumented through a birth order variable indicating the number of younger children in the household when the child was 6 years old. The second group of instrument indicates whether one of the biological parents did not live or stopped living in the household and whether a stepparent joined the household during early childhood, before the child turned 7.

The IV regression model indicates a large downward bias in the estimates for socioemotional skills, academic achievement and reading test score. The estimates on the mathematics test score and health are less clear. The downward bias of the OLS estimates suggests that parents compensate adverse shocks to child development through higher levels of investment. The results are consistent with the results from Cunha et al. (2010), Attanasio, Meghir and Nix (2015) and Attanasio et al. (2015). However, caution is needed when interpreting the results because the identification problem is not fully solved. I argue that conditional on the socio-economic status variables, the instruments affect skills only through investment. But under certain conditions, the assumption might not hold. Moreover, even though the model is identified, the instruments are still rather weak and the precision of the estimated is affected.

I estimated also a different OLS specifications which also included other forms of investment and school fixed effects. The results suggest that time investment are more important for the development of socio-emotional skills while material investments (number of books in the house, whether the child has his own books, desk room, internet at home etc.) are more important for cognitive skills. This adds to a new area of research investigating which type of investments are more productive at improving certain types of skills. My results are in line with Attanasio et al. (2015). I also find that early health is strongly associated with socio-emotional and cognitive skills and health. Early skills appear to foster the development of future skills.

Finally, this thesis adds also to the literature on ethnical minorities. There is no difference between Roma and non-Roma students in terms of socio-emotional skills after controlling for socio-economic status, early investment and health. Moreover, there is no difference in terms of investment between Roma and non-Roma parents.

My research could be extended in various way. On one hand, the effect of early childhood parenting on outcomes during adulthood (college, job market outcomes etc.) could be investigated. On the other hand, retrospective measures of early childhood socio-emotional skills or cognitive skills (less feasible) could be collected if another survey round would be implemented.

Lastly, I will like to emphasize an important area of research which still requires extensive investigation. Little is known about what determines parental behavior. One important questions is why poor parents do not invest more. Potential explanations are lack of information, wrong beliefs regarding the return to investment or poverty induced inattention. Understanding these issues is crucial in order to design interventions that are sustainable in time through improved levels of parenting.

References

Almond, Douglas and Janet Currie. 2011. In Ashenfelter, O. and Card, D. (eds), *Handbook of Labor Economics*, 4(B): 1315–1486. Elsevier, Amsterdam.

Almond, Douglas and Bhashkar Mazumder. 2013. "Fetal origins and parental responses". *The Annual Review of Economics*, 5(3):1-20.

Almlund, Mathilde, Angela Duckworth, James Heckman, and Tim Kautz. 2011."Personality Psychology and Economics". In: Hanushek, Eric, Wössmann, Ludger, and Stephen Machin (eds.): *Handbook of the Economics of Education*, Amsterdam: Elsevier, 4:1-181.

Anderson, Michael L. 2008. "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects". *Journal of the American statistical Association*. 103(484):1481-1495.

Attanasio, Orazio, Sarah Cattan, Emla Fitzsimons, Costas Meghir and Marta Rubio-Codina. 2015. "Estimating the Production Function for Human Capital: Results from a Randomized Control Trial in Colombia". *IZA Discussion Papers 8856, Institute for the Study of Labor (IZA)*

Attanasio, Orazio. 2015. "The determinants of human capital formation during the early years of life: theory, measurement and policies". *Journal of the European Economic Association*, 13(6):949–997.

Attanasio, Orazio, Costas Meghir, and Emily Nix. 2015. "Human Capital Development and Parental Investment in India," *NBER Working Papers 21740, National Bureau of Economic Research, Inc.*

Barker, D. J. 1995. "Fetal origins of coronary heart disease". *British Medical Journal*, 311(6998):171-174.

Becker, Gary S., and Nigel Tomes. 1986. "Human capital and the rise and fall of families". *Journal of labor Economics*, 4:1-39.

Behrman, Jere. R., Maria C. Calderon, Samuel H. Preston, John Hoddinott, Reynaldo Martorell, and Aryeh D. Stein. 2009."Nutritional supplementation in girls influences the growth of their children: prospective study in Guatemala". *The American Journal of Clinical Nutrition,* 90(5):1372-1379. Campbell, Fraces, Craig Ramey, Elizabeth Pungello, Joseph Sparling, and Shari Miller-Johnson. 2002. "Early childhood education: Young adult outcomes from the abecedarian project". *Applied Developmental Science*, 6(1)42-57.

Campbell, Frances, Gabriella Conti, James Heckman, Seong Moon, Rodrigo Pinto, Elizabeth Pungello, and Yi Pan. 2014."Early childhood investments substantially boost adult health". *Science* 343(6178):1478-1485.

Carneiro, Pablo, Flavio Cunha, James Heckham. 2003. "Interpreting the evidence of family influence on child development". *Paper presented at the conference The Economics of Early Child Development: Lessons for Economic Policy, Federal reserve Bank, University of Minnesota.*

Cunha, Flavio, and James Heckham. 2007. "The technology of skill formation". *American Economic Review*, 97(2):31-47.

Cunha, Flavio, and James Heckham. 2008. "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation". *Journal of Human Resources*, 43(4):738-782.

Cunha, Flavio, James Heckman, and Susanne Schennach. 2010. "Estimating the technology of cognitive and non-cognitive skill formation". *Econometrica*, 78(3):883-931.
Currie, Janet and Duncan Thomas. 1999. "Does Head Start help Hispanic children?".

Journal of Public Economics, 74(2): 235-262.

Downey, Ronald G., and Craig V. King. 1998. "Missing Data in Likert Ratings: A Comparison of Replacement Methods". *The Journal of General Psychology*, 125(2):175-191.

Engle, Patrice L., Lia C. H. Fernald, Harold Alderman, Jere Behrman, Chloe O'Gara, Aisha Yousafzai, Meena C. de Mello, Melissa Hidrobo, Nurper Ulkuer, Ilgi Ertem, Selim Iltus, and the Global Child Development Steering Group.2011."Strategies for reducing inequalities and improving developmental outcomes for young children in lowincome and middle-income countries". *Lance*, 378:1339-1353.

Evans, Gary W. 2004. "The environment of childhood poverty". *American psychologist*, 59(2):77-92.

Gertler, Paul., James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeerch, Susan Walker, Susan Chang, and Sally Grantham-McGregor. 2014. "Labor market returns to an early childhood stimulation intervention in Jamaica". *Science*, 344(6187):998-1001. Harter, Susan. 1982."The Perceived Competence Scale for Children." *Child Development* 53(1): 87-97.

Heckman, James, Jora Stixrud, and Sergio Urzua. 2006."The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior". *Journal of Labor Economics*, 24(3):411-482.

Heckman, James, Rodrigo Pinto, and Peter Savelyev. 2013."Understanding the mechanisms through which an influential early childhood program boosted adult outcomes". *American Economic Review*, 103(5):2052-2086.

Heckman, James, Seong Moon, Rodrigo Pinto, Peter Savlyev, and Adam Yavitz. 2010. "Analyzing social experiments as implemented: a reexamination of the evidence from the highscope perry preschool programs". *Quantitative Economics*, 1(1):1-46.

Heckman, James. 2007."The economics, technology, and neuroscience of human capability formation". *Proc. Natl. Acad. Sci.*, 104:13250-13255.

Helmers, C., Patnam, M., 2011. "The formation and evolution of childhood skill acquisition: Evidence from India". *Journal of Development Economics*, 95:252-266.

Howard, Kimberly S., and Jeanne Brooks-Gunn. 2009. "The role of home-visiting programs in preventing child abuse and neglect." *The future of Children,* 19(2):119-146.

Kertesi, Gabor, and Gabor Kezdi. 2011. "The Roma/Non-Roma Test Score Gap in Hungary" *American Economic Review, American Economic Association*, 101(3):519-25 Kertesi, Gabor, and Gabor Kezdi. 2014. "On the test score gap between Roma and non-Roma students in Hungary and its potential causes," *Budapest Working Papers on the Labour Market 1401, Institute of Economics, Centre for Economic and Regional Studies,*

Hungarian Academy of Sciences

Kramer, Michael S., Beverley Chalmers, Ellen D. Hodnett, Zinaida Sevkovskaya, Irina Dzikovich, Stanely Shapiro, Jean-Paul Collet, Irina Vanilovich, Irina Mezen, Thierry Ducruet, George Shishko, Vyacheslav Zubovich, Dimitri Mknuik, Elena Gluchanina, Viktor Dombrovskiy, Anatoly Ustinovitch, Tamara Kot, Natalia Bogdanovich, Lydia Ovchinikova, and Elisabet Helsing. 2001. "Promotion of breastfeeding intervention trial (probit): A randomized trial in the republic of Belarus". *Journal of the American Medical Association*, 254(4):413-420.

Maluccio, John, John Hoddinott, Jere Behrman, Reynaldo Martorell, Agnes Quisumbing, and A. Stein. 2009. "The impact of improving nutrition during early childhood on education among Guatemalan adults". *The Economic Journal*. 119(537):734-763.

Mani, Anandi, Senthil Mullainathan, Eldar Shafir, and Jiaying Zhao. 2013. "Poverty impedes cognitive function". *Science*, 341(6149):976-980.

McLoyd, Vonnie C. 1998. "Socioeconomic disadvantage and child development". *American psychologist*, 53(2):185-204.

Olds, David, Joann Robinson, LisaPettitt, Dennis Luckey, John Holmberg, Rosanna Ng, Kathy Isacks, and Karen S. C. H. Jr. 2010."Effects of home visits by paraprofessionals and by nurses: Age four follow-up results of a randomized trial.

Pediatrics". 114(6):1560-1568.

Rosenberg, Morris. 1965."Society and the adolescent self-image. Princeton". *NJ: Princeton University Press.*

Rotter, Julian. B. 1966."Generalized expectancies for internal versus external control of reinforcement". *Psychological Monographs: General & Applied*. 80(1):1-28.

Rubio-Codina, Marta, Orazio Attanasio, Costas Meghir, Natalia Varela, and Sally Grantham-McGregor. 2015."The socio-economic gradient of child development: crosssectional evidence from children 6-42 months in Bogota". *Journal of Human Resources*. 50(2):464-483.

Shah, Anuj K., Sendhil Mullainathan, and Eldar Shafir. 2012. "Some consequences of having too little". *Science*, 338(6107):682-685.

Shrive Fiona M., Heather Stuart, Hude Quan and William A. Ghali. 2006. "Dealing with missing data in a multi-question depression scale: a comparison of imputation methods". *BMC Medical Research Methodology*, 13:6-57.

Simonovits, Gabor, and Gabor Kezdi. 2016. "Economic Hardship Triggers Identification with Disadvantaged Minorities". *The Journal of Politics*, 78(3).

Thompson, R., Nelson, C. 2001. "Developmental science and the media: early brain development". *American Psychology*, 56:5-15.

Todd Petra E. and Kenneth I. Wolpin. 2003."On The Specification and Estimation of The Production Function for Cognitive Achievement". *Economic Journal, Royal Economic Society*, 113(485):F3-F33.

Victora, Cesar G., Linda Adair, Caroline Fall, Pedro C. Hallal, Reynaldo Martorell, Linda Richter, and Harshpal S. Sachdev. 2008. "Maternal and child undernutrition: consequences for adult health and human capital". *Lancet*, 371(9609):340-357. Walker, Susan P., Theodore D. Wachs, Julie Meeks Gardner, Betsy Lozoff, Gail A.
Wasserman, Ernesto Pollitt, and Julie A. Carter. 2007. "Child development: risk factors for adverse outcomes in developing countries". *The Lancet*, 369(9556):145-157.
Walker, Susan, Theodore Wachs, Susan Grantham-McGregor, Maureen Black, Charles Nelson, Sandra Huffman, Helen Henningham, Susan Chang, Jena Hamadani, Betsy Lozoff, Julie Gardner, Christine Powell, Atif Rahman, and Linda Richter. 2011. "Inequality in early childhood: risk and protective factors

for early child development". Lancet, 378(9799):1325-1338.

Appendix

Table A1. Rosenberg's Self-Esteem Scale – strongly disagree, disagree, agree, strongly	
agree – coded 0-3 – higher score reflects higher level of self-esteem	

	Age 15		Age	18
Self Esteem – main sample	Mean SD		Mean	SD
I feel that I'm a person of worth, at least on an equal plane with others.	2.05	0.58	2.12	0.67
I feel that I have a number of good qualities.	2.12	0.57	2.27	0.59
I am able to do things as well as most other people.	2.26	0.6	2.39	0.61
I feel like I have a lot of things to be proud of	0.88	0.79	0.77	0.82
I take a positive attitude toward myself.	2.07	0.61	2.09	0.67
On the whole, I am satisfied with myself.	2.06	0.64	2.08	0.7
All things considered, I tend to untalented, unsuccessful people	2.22	0.75	2.22	0.84
I wish I could have more respect for myself.	1.85	0.83	2	0.89
I certainly feel useless at times.	2.09	0.8	2.09	0.9
Sometimes I think I'm good at nothing	2.13	0.81	2.2	0.88
XX7 * 1 / 11 11 * 1 /				

Weighted by sampling weights

Table A2. Rotter's External Locus of Control

	Age 15		Age 16	
External Locus of Control – main sample	Mean	SD	Mean	SD
0-What happens to me is my own doing				
1-Sometimes I feel that I don't have enough control over the direction my	0.18	0.38	0.2	0.4
life is taking -1				
0-When I make plans, I am almost certain that I can make them work.				
1-It is not always wise to plan too far ahead because many things turn out to-	0.31	0.46	0.3	0.46
be a matter of good or bad fortune anyhow.				
0-In my case getting what I want has little or nothing to do with luck.	0.10	0.4	0.21	0.41
1-Many times we might just as well decide what to do by flipping a coin.	0.19			0.41
0-It is impossible for me to believe that chance or luck plays an important				
role in my life.	0.27	0.49	0.27	0.49
1-Many times I feel that I have little influence over the things that happen to	0.57	0.48	0.57	0.48
me.				

Weighted by sampling weights

Social Competences – main sample		Age 15		a
	Mean	SD	Mean	SD
a) Few Friends/ b) Many Friends	2.48	0.72	2.43	0.74
a) Not really popular among others/b) Popular among others	2.06	0.78	2.15	0.73
a) Difficult to be liked/b)Easy to be liked	1.89	0.92	2.33	0.77
a) Often alone/b) Often spending time with other young people	2.31	0.79	2.44	0.81
a) Difficult to make friends/b) Easy to make friends	2.05	0.97		
a) Opinion often ignored by others/b) Opinion important for others	2.18	0.75		
a) Others mostly dislike me/b) Others mostly like me	2.31	0.67	2.52	0.62

Table A3. Harter's Social Competences - 0-completely agree a); 1-rather agree a); 2-rather agree b); 3-completely agree b)

Weighted by sampling weights

Table A4. Depression at age 15

Depression Age 15 – main sample	Mean	SD
0 - I'm rarely sad; 1 - I'm often sad; 2 - I'm always sad	0.1	0.32
0 - Many things bring me joy; 1 - Some things bring me joy; 2 - Nothing brings me joy	0.21	0.41
0 - I get along well with others; 1 - Often argue with others; 2 - Almost always arguing	0.1	0.33
0 - I never thought of killing myself; 2 - I thought of killing myself but would never do it; 2 - I would like to kill myself	0.12	0.35
0 - I love myself; 1 - I don't like myself; 2 - I hate myself	0.15	0.36
0 - At times, I feel like crying; 1 - Often I feel like crying; 2 - I always feel like crying	0.06	0.26
0 - It's not difficult for me to decide; 1 - Sometimes it is difficult to decide; 2 - I can't decide on anything	0.36	0.5
0 - I'm sure someone likes me; 1 - I'm not sure somebody likes me; 2 - Nobody likes me	0.07	0.27

Weighted by sampling weights

Table A5. Expectations - coded probabilities: 0=0-24%; 1=25-49%; 2=50-74% 3=75-100%

	Age	Age 17		21
Expectations – main sample	Mean	SD	Mean	SD
The probability that:				
Satisfied with her/his life in 3 years	2.28	0.75	2.43	0.72
To have good health in 3 years	2.51	0.7	2.62	0.61
Exercise regularly at least 4 times a week in 3 years	1.76	1.13	1.86	1.16
Will be the victim of a serious accident in the next 3 years	2.28	0.94	2.28	0.94
Her/his mother will live to be at least 80 years old	2.47	0.78	2.52	0.75
Will smoke regularly in 3 years from now	2.37	1.08	2.29	1.15
Will smoke regularly 20 years from now	2.48	0.96	2.48	0.97
Will graduate high-school?	2.51	0.9	1.19	1.24
Will acquire a higher education diploma	1.59	1.24	1.48	1.36
Will earn more money than the average person at age 35	1.86	0.92	1.65	1.09
Will be in the top 10% of individuals with highest earnings	0.76	0.95	0.55	0.89
After leaving school will find regular employment	2.35	0.77	2.26	0.86
Will look for jobs paying more than 100,000 HUF/month after taxes	2.08	0.94	2.02	1.11

Weighted by sampling weights

Table A6. Learning age 16; coded - 0 almost never 1 sometimes 2 often 3 always

Learning Age 16 – main sample	Mean	SD
I always start studying by thinking through what do I have to learn exactly.	2.04	0.92
While studying I try to check whether I remember what I have already learned.	1.9	0.89
While studying I try to think what concepts I still don't understand.	1.99	0.87
While studying I check whether I remember the most important things.	1.97	0.85
If I don't understand something I try to look it up somewhere else.	1.71	0.94
While studying I try to remember everything related to the subject\topic.	2.14	0.78
While learning more things I try to take notes	2.29	0.73
While studying I try to remember as many things as possible.	2.1	0.86
I always try to learn the new material so I could repeat it.	1.69	0.98
While studying I try to tell myself the material again and again.	1.59	0.92
While studying I try to connect the new material to my knowledge from other courses.	1.4	0.94
While studying I try to think about how I can use this knowledge in real life.	1.83	0.85
While studying I try to understand the material by connecting it to my existing knowledge.	1.85	0.86
Weighted by sampling weights		

Table A7. Early childhood parental investment items

Early Childhood Investment – main sample		
Answers of parents		
When the child was at preschool age how often the parents:		
0 Never or almost never; 1 Few times a month 2 Once or twice a week; 3 3-5 times a week; 4 Every day or almost every day	Mean	SD
Told stories from storybook or by heart	3.11	1.09
Viewed a story on TV with the child?	3.47	0.93
Listened to a radio story with the child?	0.42	0.94
Listened to a story on CD, disk or tape with the child?	1.01	1.3
Answers of children		
When you were 3-4 years old how often did the parents:		
0 Less often than twice a year; 1 Twice a year; 2 Every two to three months; 3 Once or twice a month; 4 Weekly; 5 Several times a week	Mean	SD
Read a story to you before going to bed	4.12	1.54
Played a board game with you	3.56	1.41
Cooked, made gifts, drawn with you	3.03	1.59
Went to concerts, puppetry, theater with you	1.32	1.37
Went on trips or played sports with you	1.79	1.49
Weighted by sampling weights		

Name	Description						
	Standardized Index reflecting frequency of parenting interaction with the child						
Early Childhood Investment	at preschool age (parents and children retrospective answers)						
	Health, Gender, Test scores						
Body Height	Standardized body height according to age and gender						
Low weight at birth	Was the child's weight less than 2500 grams at birth? (yes/no)						
Child Health - good or excellent	Self-evaluation of child's health when he/she was 15 (four ittem scale - poor, fair, good, excellent) (yes/no)						
Early Childhood Health Issues	Dummies for whether suffered (before age of 6) from issues with: speech, hearing, circulation, neurological, gastrointestinal, asthma, respiratory, mental, emaciation, obesity or other						
Gender - girl/boy	Gender of the child? (1 girl 0 boy)						
Math test score	Standardized mathematics score						
Reading test score	Standardized reading score						
	Home/Parenting						
Nursery	Did the child attend nursery school? (yes/no)						
Home Score - emotional subscale	Standardized index for the Home emotional subscale items						
subscale	Standardized index for the Home cognitive subscale items						
Computer	Does the child have his own computer? (yes/no)						
Child's Books	Does the child own any books that are not textbooks? (yes/no)						
Desk	Does the child have his own desk? (yes/no)						
Room	Does the child have his own room? (yes/no)						
Caregiver's educational aspirations	Highest level of education would want the child to complete (0 elementary school 1 secondary/high school/professional training 2 higher education)						
Books in the household	Number of books in the household (<50/50-150/150-300/300-600/600-1000/>1000)						
Internet connection at home	Is there internet connection in the household?						
Abused before age 15	Was the child abused psychologically, physically or sexually before the age of 15? (yes/no) (child answers at age 21)						
Abused after age 15	Was the child abused psychologically, physically or sexually after the age of $152 (ves/po)$ (child answers at age 21)						
Working for the family at	13: (yes/ho) (clifid answers at age 21)						
age 15	Did the child use to work for the family when he was 15? (yes/no)						
Worked for the family	Did the child use to work for the family when he was at primary school age?						
primary senioor age	Socio-economic status						
Roma Ethnicity	Either of the biological parents stating being Roma as first or second nationality in the first two waves of the survey. Or the child declaring to be Roma as first or second nationality in the 4th or 6th waves.						
Mother education	Highest level of education completed by the mother (0-8 elementary school, vocational school, high school, high education)						
Father education	Highest level of education completed by the father (0-8 elementary school, vocational school, high school, high education)						
employment Father's current	Whether mother was employed in the first wave (yes/no)						
employment	Whether father was employed in the first wave (yes/no)						

Table A8. Description of the independent variables

Father's health- excellent	Father's subjective health during the first wave of the survey (four item scale - poor, fair, good, excellent) (yes/no)
Mother's Health - excellent	Mother's subjective health during the first wave of the survey (four item scale - poor, fair, good, excellent) (yes/no)
income	The logarithm of the household's income during the first wave of the survey
Logarithm of household size	The logarithm of the number of household members during the first wave of the survey
Bathroom	Whether there is a bathroom in the house (yes/no)
Car	Does the family own a car?
Apartment size per capita	Total size of the apartment (in square meters) divided by the number of people in the household
person	Number of rooms divided by the number of people in the household
regular work Number of unemployed	Number of people with regular work in the household
people	Number of unemployed people in the household
Number of old age people	Number of old age people in the household
Number of pensioners Younger children when	Number of pensioners in the household
child was 6 Older children when child	Number of younger children in the household when child was 6 years old
was 6	Number of older children in the household when child was 6 years old
Number of siblings	Number of siblings when the child was 15
Poverty - food	Was there ever not enough money for food in the past 12 months? (yes/no)
Poverty - rent	Was there ever not enough money for rent in the past 12 months? (yes/no)
Poverty - heating	Was there ever not enough money for heating in the past 12 months? (yes/no)
Social assistance free meals Social assistance discounted	Whether child receives free meals at schools? (yes/no)
meals Social assistance free	Whether child receives discounted meals at schools? (yes/no)
textbooks Social assistance education	Whether child receives free textbooks? (yes/no)
aid	Whether child receives other forms of education aid? (yes/no)
Childcare benefits - mother Unemployment benefits -	Whether the mother receives childcare benefits? (yes/no)
mother	Whether the mother receives unemployment benefits? (yes/no)
Pension - mother Other social benefits -	Whether the mother is a pensioner? (yes/no)
mother	Whether the mother is receiving other forms of social benefits (yes/no)
Childcare benefits - father Unemployment benefits -	Whether the father receives childcare benefits? (yes/no)
father	Whether the father receives unemployment benefits? (yes/no)
Pension - father	Whether the father is a pensioner? (yes/no)
Other social benefits - father	Whether the father is receiving other forms of social benefits (yes/no)
Neighborhood Index	Standardized index reflecting the living standards and safety in the neighborhood (interviewer and responder answers)
Region	The region of residence
Place of residence type	Place of residence type - rural, Budapest, large cities, other cities

	Academic Achievement				
Early Childhood Investment		Socio-Emotional Skills Age	Socio-Emotional Skills Age 15		
Bedtime stories	0.68	Self Esteem Age 15	0.67	Math	0.77
Games	0.94	External Locus of Control Age 15	-0.40	Language	0.88
Cooking, crafting	0.61	Social Competences Age 15	0.43	Literature	0.92
Concerts, theater	0.52	Depression Age 15	-0.70	Foreign Language	0.77
Trips, sports	0.53			Year Average	0.87
Story telling (parent's answer)	0.50			Reading Score	0.62

Table A9. Exploratory Factor Analysis. Factor Loadings

Table A10. Summary statistics for dependent variables and main explanatory variable

	Main Sample		Full Sample			
	Mean	SD	Obs	Mean	SD	Obs
Socio-Emotional Skills Age 15	0.05	0.96	5231	0	1	9866
Life Satisfaction Age 15	0.05	0.96	5231	0	1	9993
Satisfied with Life Age 21	0.87	0.34	5231	0.86	0.35	6955
Self Esteem Age 15	0.02	0.99	5231	0	1	9986
Self Esteem Age 18	0.01	1.01	5231	0	1	7766
Locus of Control Age 15	-0.05	0.99	5231	0	1	9979
Locus of Control Age 16	-0.03	0.99	5231	0	1	7607
Social Competences Age 15	0.03	0.98	5231	0	1	9895
Social Competences Age 17	0.04	0.98	5231	0	1	8589
Expectations Age 17	0.07	0.97	5231	0	1	8438
Expectations Age 21	0.06	0.98	5231	0	1	6560
Learning Age 16	0.03	0.98	5231	0	1	8660
Depression Age 15	-0.06	0.94	5231	0	1	9998
Academic Achievement Age 15	0.1	0.98	5231	0	1	9733
Math Score ^a	0.04	1.04	4911	-0.05	1.04	9167
Reading Test Score ^a	-0.01	1.01	5231	-0.11	1.03	10011
Math Average	0.09	1	5231	0	1	9960
Language Average	0.09	0.99	5231	0	1	9958
Literature Average	0.1	0.97	5231	0	1	9971
Conduct Average	0.1	0.95	5231	0	1	9961
Foreign Language Average	0.11	0.96	5231	0	1	9961
Year School Average	0.09	0.99	5231	0	1	9806
Health - Excellent	0.38	0.49	5193	0.38	0.49	9919
EC Investment	0.05	0.97	5231	0	1	9902

Weighted by sampling weights ^a Standardized using NABC test score population mean

	Main Sample		Full Sample			
	Mean	SD	Obs	Mean	SD	Obs
Gender - Girl	0.48	0.5	5231	0.49	0.5	10011
Low Birth Weight	0.07	0.25	5192	0.08	0.27	9869
Height for Age Score	0.04	1.09	5223	0.02	1.06	9998
Early Childhood Health Issues						
Speech	0.18	0.38	5229	0.17	0.37	9982
Hearing	0.01	0.1	5231	0.01	0.1	10011
Muscular	0.02	0.15	5231	0.02	0.15	10009
Circulatory	0.01	0.12	5231	0.01	0.12	10009
Neurological	0	0.07	5231	0.01	0.08	10009
Gastrointestinal	0.01	0.08	5231	0.01	0.08	10009
Asthma	0.05	0.21	5230	0.05	0.22	10009
Respiratory	0.03	0.16	5230	0.03	0.17	10009
Mental	0.09	0.28	5229	0.09	0.28	10008
Emaciation	0	0.05	5231	0	0.05	10009
Obesity	0	0.05	5231	0	0.05	10010
Other	0	0.05	5231	0	0.05	10010
Nursery School	0.2	0.4	5221	0.21	0.41	9973
Home Emotional	0	0.99	5123	0	1	9727
Home Cognitive	0.08	0.98	5178	0	1	9862
Internet	0.49	0.5	5222	0.47	0.5	9990
Books at home						
<50	0.1	0.3	5208	0.1	0.3	9939
~50	0.11	0.31	5208	0.14	0.34	9939
50-150	0.11	0.31	5208	0.12	0.32	9939
150-300	0.22	0.42	5208	0.22	0.42	9939
300-600	0.2	0.4	5208	0.19	0.39	9939
600-1000	0.17	0.38	5208	0.16	0.36	9939
>1000	0.08	0.28	5208	0.08	0.28	9939
Child's Books	0.94	0.24	5196	0.92	0.27	9937
Desk	0.9	0.3	5226	0.87	0.33	9992
Room	0.7	0.46	5226	0.7	0.46	9991
Computer	0.42	0.49	5224	0.41	0.49	9986
Educational Aspirations						
Elementary	0.08	0.27	5176	0.1	0.3	9841
Secondary	0.27	0.44	5176	0.29	0.46	9841
Higher	0.65	0.48	5176	0.6	0.49	9841
Abused before Age 15	0.41	0.49	5060	0.41	0.49	6622
Abused after Age 15	0.25	0.43	4987	0.26	0.44	6501
Working for Family Age 15	0.04	0.2	5220	0.04	0.2	9983
Worked for Family Primary School Age	0.06	0.24	5210	0.06	0.23	9957

Table A11. Summary Statistics for gender, early health and home environment

Weighted by sampling weights

	Main Sample		Full Samp		le	
	Mean	SD	Obs	Mean	SD	Obs
Roma	0.08	0.27	5231	0.09	0.29	10011
Mother's Education Level						
Higher	0.22	0.41	5136	0.2	0.4	9771
0-8th grades	0.18	0.38	5136	0.21	0.41	9771
Vocational	0.26	0.44	5136	0.25	0.43	9771
Secondary	0.35	0.48	5136	0.34	0.47	9771
Father's Education Level						
Higher	0.17	0.37	4257	0.16	0.37	7963
0-8th grades	0.12	0.33	4257	0.15	0.35	7963
Vocational	0.45	0.5	4257	0.45	0.5	7963
Secondary	0.26	0.44	4257	0.25	0.43	7963
Mother Employed	0.7	0.46	5139	0.67	0.47	9772
Father Employed	0.81	0.39	4239	0.79	0.41	7925
Father Excellent Health	0.12	0.32	5231	0.11	0.31	10011
Mother Excellent Health	0.12	0.32	5231	0.12	0.32	10011
Logarithm of Income	11.86	0.45	4615	11.83	0.48	8626
Logarithm of Household Size	1.4	0.29	5231	1.4	0.31	10011
Bathroom	0.96	0.19	5224	0.95	0.21	9986
Car	0.67	0.47	5215	0.64	0.48	9967
Computer	0.84	0.36	5219	0.82	0.39	9975
Apartment Size/Person	23.45	9.9	5157	23.11	10.49	9837
Number of Rooms/Person	0.78	0.28	5213	0.77	0.3	9963
Number of Household Members with:						
Regular Work	1.45	0.78	5231	1.39	0.81	10011
Unemployed	0.18	0.47	5231	0.2	0.5	10011
Old Age	0.1	0.34	5231	0.1	0.35	10011
Pension	0.24	0.51	5231	0.25	0.53	10011
Neighborhood Index	0.07	0.96	5103	0	1	9772
Number of Siblings	1.32	1.07	5231	1.34	1.17	10011
Poverty - Food	0.05	0.22	5231	0.06	0.24	10011
Poverty - Rent	0.11	0.32	5231	0.14	0.34	10011
Poverty - Heating	0.1	0.3	5231	0.11	0.31	10011
Social Assistance						
Free Meals at Schools	0.08	0.27	5224	0.09	0.28	9990
Discounted Meals	0.26	0.44	5222	0.26	0.44	9986
Free Textbooks	0.57	0.5	5225	0.59	0.49	9994
Education Aid	0.24	0.43	5212	0.26	0.44	9973
Social Benefits						
Mother Childcare	0.51	0.5	5231	0.51	0.5	10011
Mother Unemployment	0.04	0.19	5231	0.04	0.2	10011
Mother Pension	0.07	0.25	5231	0.08	0.26	10011

Table A12. Summary statistics for socio-economic status and instrumental variables

Mother Social	0.06	0.23	5231	0.06	0.24	10011
Father Childcare	0.25	0.43	5231	0.22	0.42	10011
Father Unemployment	0.03	0.16	5231	0.03	0.17	10011
Father Pension	0.07	0.25	5231	0.07	0.25	10011
Father Social	0.02	0.15	5231	0.03	0.17	10011
Instrumental Variables						
Children 0-6 years old in household when the child was 6	0.84	0.84	5231	0.84	0.89	10011
Children 7-10 years old in household when the child was 6	1.02	0.66	5231	1.04	0.66	10011
Mother employed when the child was 3-6 years old	0.75	0.43	4748	0.75	0.43	8900
Father employed when the child was 3-6 years old	0.98	0.14	3790	0.97	0.16	6910
Separated from mother age 0-7	0.02	0.12	5231	0.02	0.15	10011
Separated from father age 0-7	0.14	0.34	5231	0.16	0.36	10011
Stepfather age 0-7	0.03	0.16	5231	0.03	0.18	10011
Stepmother age 0-7	0.01	0.08	5231	0.01	0.10	10011

Weighted by sampling weights

Figure A 1. Histograms of the main dependent and explanatory variables. Weighted by sampling weights





	(1)	(2) Socio- Economic Status	(3) Early Childhood Health	(4) Gender	(5) Home Environment	(6) School FE	(7) Class FE
Socio	0 210***	0 168***	0 167***	0 170***	0 136***	0 136***	0 120***
emotional Skills;	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.016)
N=9,773							
Life	0.204***	0.089***	0.089***	0.089***	0.059***	0.062***	0.071***
Satisfaction Age 15; N=9 892	(0.013)	(0.012)	(0.012)	(0.012)	(0.013)	(0.013)	(0.016)
Life	0 039***	0.011*	0.011*	0.011*	0.003	0.004	0.004
Satisfaction Age 21; N=6 893	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.008)
Self Esteem	0.168***	0.144***	0.145***	0.148***	0.124***	0.122***	0.105***
Age 15; N=9,889	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.016)
Self Esteem	0.094***	0.073***	0.074***	0.077***	0.056***	0.052***	0.066***
Age 18; N=7,693	(0.015)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.019)
External Locus of	-0.117***	-0.074***	-0.074***	-0.074***	-0.045***	-0.050***	-0.059***
Control Age 15; N=9,884	(0.009)	(0.011)	(0.011)	(0.011)	(0.012)	(0.012)	(0.014)
External Locus of	-0.103***	-0.071***	-0.072***	-0.072***	-0.053***	-0.054***	-0.046**
Control Age 16; N=8,501	(0.015)	(0.016)	(0.017)	(0.016)	(0.016)	(0.017)	(0.021)
Social Competences	0.141***	0.116***	0.116***	0.116***	0.094***	0.092***	0.084***
Age 15; N=9,797	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.016)
Social Competences	0.082***	0.069***	0.068***	0.070***	0.059***	0.064***	0.079***
Age 17; N=8,501	(0.014)	(0.015)	(0.016)	(0.016)	(0.015)	(0.015)	(0.016)
Expectations	0.281***	0.126***	0.125***	0.125***	0.076***	0.072***	0.069***
Age 17; N=8,352	(0.021)	(0.019)	(0.019)	(0.019)	(0.017)	(0.018)	(0.019)
Expectations	0.280***	0.092***	0.089***	0.089***	0.047***	0.041***	0.059***
Age 21; N=6,503	(0.019)	(0.016)	(0.016)	(0.016)	(0.015)	(0.015)	(0.019)
Learning Age	0.181***	0.112***	0.112***	0.106***	0.059***	0.055***	0.052***
16; N=8,575	(0.016)	(0.015)	(0.015)	(0.014)	(0.014)	(0.014)	(0.016)
Depression	-0.177***	-0.124***	-0.121***	-0.125***	-0.097***	-0.099***	-0.099***
Age 15; N=9,897	(0.013)	(0.014)	(0.014)	(0.013)	(0.014)	(0.014)	(0.019)
Academic	0.247***	0.091***	0.089***	0.080***	0.020**	0.022**	0.022*
Achievement Age 15; N=9.636	(0.014)	(0.012)	(0.011)	(0.011)	(0.010)	(0.010)	(0.012)
Math Score	0.237***	0.052***	0.049***	0.051***	0.009	0.008	0.018
Age 15; N=9,083	(0.016)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)
Reading	0.258***	0.076***	0.073***	0.066***	0.017	0.017	0.018*
Score Age 15; N=9,902	(0.015)	(0.011)	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)
Health Age	0.051***	0.026***	0.028***	0.029***	0.021***	0.023***	0.025***
15 -	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)

Table A13. Robustness Check. OLS Regression results for the full available sample

*** p<0.01, ** p<0.05, * p<0.1.Clustered standard errors at school level in brackets. Weighted by sampling weights.

Table A 14. Regression results using the main sample. Specification in columns with (1) control for socio-economic status and early childhood health while columns with (2) control in addition for home environment and school fixed effects

	(1)	(2)	(1)	(2)
	Socio- Emotional Skills	Socio- Emotional Skills	Health – excellent age 15	Health – excellent age 15
Fouly shildhood investment	0 159***	0 101***	0 0 0 0 * * *	0.010*
Early childhood investment	0.158***	0.121****	0.028****	0.019*
Mother education - baseline high	0.030	0.049	0.115	0.080
0-8 grade		0.021	_0 1/18***	_0 11/**
Vocational	-0.074	0.021	-0.148	-0.114
Secondary	0.016	0.045	-0.078	-0.049
Missing mother education	-0.036	0.199	0.151	0.220
Father education - baseline high	er education	0.177	0.131	0.220
0-8 grade	-0.005	0.053	-0.031	0.004
Vocational	-0.057	-0.031	-0.037	-0.012
Secondary	-0.051	-0.044	-0.028	-0.012
Missing father education	0.179	0.133	-0.118	-0.140
Mother employed	-0.016	-0.049	-0.033	-0.028
Missing mother employed	0.041	0.055	-0.060	-0.053
Father employed	-0.000	-0.039	-0.013	-0.026
Missing father employed	-0.372**	-0.214	-0.020	0.048
Father excellent health	0.034	0.021	0.088***	0.079***
Missing father health	0.040	-0.026	0.143*	0.099
Mother excellent health	0.019	0.034	0.140***	0.139***
Missing mother health	0.005	-0.090	-0.086	-0.131*
Logarithm of income	0.041	0.048	0.027	0.027
Missing income	-40.030	-47.438	-26.605	-26.793
Logarithm of HH size	-0.016	-0.023	-0.046	-0.029
Bathroom in the house	-0.068	-0.101	-0.016	-0.046
Missing bathroom	-0.543	-0.725*	-0.263	-0.450*
Car	0.083*	0.045	-0.025	-0.027
Missing car	0.691*	0.634	-0.229	-0.255
Computer	0.006	-0.016	0.017	0.008
Missing computer	0.132	0.277	0.352**	0.458***
Apartment size/person	0.001	-0.000	0.002	0.002*
Missing apartment size	-1.037	0.182	-1.808	-2.095*
Rooms/person	0.000	0.000	-0.000	-0.000
People in HH with:				
Regular work	0.015	0.036	0.010	0.013
Unemployed	-0.029	-0.011	-0.035*	-0.029
Old age	0.027	0.013	-0.000	-0.006
Pensioners	-0.019	-0.003	0.005	-0.001

Neighborhood index	0.042**	0.011	0.009	-0.002
Missing neighborhood index	-42.418**	-10.595	-8.829	2.220
Siblings	-0.009	-0.014	0.009	0.010
Poverty - Food	-0.131*	-0.106	-0.053	-0.035
Poverty - Rent	-0.148***	-0.105**	-0.006	0.003
Poverty - Heating	-0.102	-0.074	-0.034	-0.038
Missing poverty indicators	0.459	0.327	0.197	0.214
Free meals at schools	-0.126*	-0.090	0.031	0.041
Missing free meals	-0.335	-0.372	0.039	0.086
Discounted meals at school	-0.027	-0.027	0.005	0.008
Missing discounted meals	-0.289	-0.090	-0.228	-0.077
Free textbooks	-0.075*	-0.055	-0.052**	-0.051**
Missing free textbooks	0.012	-0.044	-0.171	-0.195
Education aid	0.104**	0.079*	0.043	0.038
Missing education aid	-0.055	-0.031	-0.008	0.009
Mother - social assistance				
Childcare	-0.067	-0.051	-0.017	-0.019
Unemployment benefits	0.121	0.099	-0.008	-0.003
Pension	-0.066	-0.062	-0.012	0.014
Other	0.000	0.002	0.012	0.011
Missing mother social	0.044	0.012	0.000	0.011
assistance	-0.056	-0.125	-0.056	-0.040
Father - social assistance				
Childcare	-0.067	-0.062	-0.004	-0.005
Unemployment benefits	0.099	0.066	0.060	0.064
Pension	0.023	0.002	0.009	0.013
Other	-0.187	-0.243	0.013	0.013
Missing father social assistance	0.186**	0 190**	-0.015	0.000
I ow weight at birth	-0.057	-0.049	-0.026	-0.034
Missing weight at hirt	-0.022	0.049	0.020	-0.015
Height for age score	0.000	0.013	-0.000	-0.015
Early childhood health issues	0.000	0.000	-0.000	-0.000
Speech	0 126***	0.000**	0.045***	0.024**
Missing speech	-0.130***	-0.099**	-0.043***	-0.034**
Missing speech Hearing	0.388	0.451**	-0.330***	-0.104*
Maganlatan	-0.551***	-0.203*	-0.104*	-0.093*
Musculature	0.023	0.030	-0.077***	-0.060
Circulatory	-0.096	0.007	-0.15/***	-0.129***
Neurological	-0.464	-0.429	-0.109	-0.113
Gastrointestinal	0.192	0.286**	0.107	0.124
Asthma	-0.079	-0.063	-0.068	-0.063
Missing Asthma	-0.063	0.320*	-0.095	-0.2/1***
Respiratory	-0.163**	-0.163**	-0.055	-0.046
Mental	-0.000	-0.011	-0.075***	-0.081***
Missing mental	-0.068	-0.407***	-0.427***	-0.374***
Emaciation	-0.647**	-0.591**	-0.088	-0.074
Obesity	-0.658	-0.531	-0.110	-0.096
Other	-0.518	-0.448	-0.207*	-0.237**
Gender - Girl	-0.180***	-0.237***	-0.064***	-0.070***
Nursery School		-0.003		0.037**
Missing nursery		-0.221		0.166
HOME -emotional subscale		0.076***		0.016
HOME -cognitive subscale		0.057**		0.015
Missing HOME emotional		-75.762***		-15.609
Missing HOME cognitive		-56.726**		-15.156

Internet		0.042		0.012
Missing internet		0.172		-0.062
Books at home - baseline				
>1000				
<50		0.154**		0.028
~50		0.060		0.011
50-150		0.053		-0.016
150-300		0.016		-0.030
300-600		-0.021		-0.039
600-1000		0.040		-0.031
Missing number of books		0.046		0.331***
Child own books (not		0.014		0.023
textbooks)		-0.014		0.025
Missing child's books		-0.105		-0.071
Child's desk		0.176**		0.046*
Missing child's desk		-0.145		-0.289
Child own room		-0.014		-0.018
Child own computer		-0.019		0.005
Missing child's computer		0.308		0.456*
Educational aspirations of paren	ts for the child			
- baseline elementary				
Secondary/vocational		0.174***		-0.001
Higher		0.376***		0.075***
Missing aspirations		0.443***		0.009
Abused before age 15		-0.111**		-0.041**
Missing abused		0.116		-0.031
Abused after age 15		-0.168***		-0.041*
Missing abused after age 15		-0.083		-0.012
Working for family age 15		0.126		0.058
Missing working for family age		0.250		0.027
15		-0.230		0.037
Working for family primary		-0 251***		-0.030
school age		0.201		0.050
Missing working for family		-0.250		-0.224**
primary school age				
Dequered	0.102	0.174	0.082	0.122

R-squared0.1020.1740.0830.133*** p<0.01, ** p<0.05, * p<0.1.Clustered standard errors at school level in brackets. Weighted by sampling weights.</td>

Table A 15. Regression results using the main sample. Specification in columns with (1) control for socio-economic status and early childhood health while columns with (2) control in addition for home environment and school fixed effects

	(1)	(2)	(1)	(2)	(1)	(2)
VADIADIES	Academic	Academic	Reading	Reading	Math Same	Math Saama
VARIABLES	achievement	achievement	Score	Score	Math Score	Math Score
EC Investment	0.080***	0.010	0.057***	0.003	0.054***	-0.001
Roma	-0.092	-0.082*	-0.105*	-0.064	-0.308***	-0.256***
Mother education -baseline higher education						
0-8 grade	-0.623***	-0.272***	-0.531***	-0.148**	-0.496***	-0.119*
Vocational	-0.558***	-0.306***	-0.508***	-0.221***	-0.484***	-0.188***
Secondary	-0.250***	-0.168***	-0.248***	-0.134***	-0.191***	-0.085*
Missing mother education	-0.978***	-0.435**	-0.867***	-0.226	-1.183***	-0.595***
Father education- baseline higher education						
0-8 grade	-0.493***	-0.251***	-0.557***	-0.276***	-0.590***	-0.336***
Vocational	-0.369***	-0.199***	-0.427***	-0.213***	-0.520***	-0.297***
Secondary	-0.173***	-0.129***	-0.237***	-0.156***	-0.305***	-0.230***
Missing father education	-0.087	-0.300*	0.295	0.124	0.089	-0.088
Mother employed	0.023	-0.034	0.035	-0.021	0.067	0.028
Missing mother employed	0.189	0.038	0.352	0.227	0.861***	0.787***
Father employed	0.027	-0.009	0.046	0.017	0.040	0.001
Missing father employed	-0.178	0.145	-0.272	-0.040	-0.356**	-0.086
Father excellent health	-0.033	-0.002	-0.061	-0.028	-0.057	-0.035
Missing father health	-0.128	-0.053	-0.234	-0.148	-0.088	-0.038
Mother excellent health	-0.050	0.015	0.016	0.047	-0.080*	-0.044
Missing mother health	-0.022	-0.009	-0.100	-0.105	-0.180	-0.211
Logarithm of income	0.046	0.033	0.044	-0.001	0.007	-0.034
Missing income	-45.169	-32.636	-43.810	0.791	-7.094	33.500
Logarithm of HH size	0.118	0.197**	-0.052	-0.004	0.149*	0.170**
Bathroom in the house	0.003	-0.053	0.042	-0.095	-0.025	-0.129
Missing bathroom	0.294	-0.223	0.643	0.118	0.993***	0.388*
Car	0.039	-0.010	0.005	-0.045	0.017	-0.032
Missing car	0.477	0.492	0.266	0.327	0.272	0.429
Computer	0.157***	0.014	0.222***	0.059	0.185***	0.023
Missing computer	-0.937***	-0.528	-0.846***	-0.534**	-0.840**	-0.655***
Apartment size/person	0.002	-0.000	0.002	-0.000	0.004**	0.001
Missing apartment size	-1.657	0.553	-1.703	0.008	-3.712**	-1.465
Rooms/person	0.000	0.000	0.000	0.000	0.000	0.000
People in household with:						
Regular work	-0.127***	-0.069**	-0.068**	-0.025	-0.113***	-0.066*
Unemployed	-0.117***	-0.087***	-0.100***	-0.084***	-0.098**	-0.089**
Old age	-0.028	-0.051	-0.004	-0.009	0.025	0.013
Pensioners	-0.009	0.025	-0.038	-0.032	-0.097*	-0.068
Neighborhood index	0.066***	0.010	0.055***	0.003	0.093***	0.039*
Missing neighborhood index	-65.469***	-9.753	-55.075***	-3.442	-93.496***	-39.263*
Siblings	-0.064***	-0.061***	-0.037**	-0.033**	-0.035*	-0.042**
Poverty - Food	-0.108*	-0.025	-0.163**	-0.105*	-0.067	-0.026
Poverty - Rent	-0.065	-0.008	0.022	0.053	0.064	0.106**
Poverty - Heating	0.051	0.038	-0.013	-0.029	-0.105**	-0.111**
Missing poverty indicators	0.142	0.109	0.069	0.166	-0.032	0.021

Free meals at schools	0.010	0.047	-0.140**	-0.079*	-0.107	-0.066
Missing free meals	0.442	0.480**	0.317	0.115	0.283	0.177
Discounted meals at school	0.002	-0.011	-0.019	-0.015	0.011	0.021
Missing discounted meals	0.025	-0.032	0.053	-0.189	0.228	0.169
Free textbooks	-0.032	-0.042	-0.044	-0.046	-0.030	-0.050
Missing free textbooks	-0.013	-0.025	-0.152	-0.043	-0.117	-0.124
Education aid	-0.011	-0.030	0.077*	0.056	0.015	0.017
Missing education aid	-0.071	0.161	-0.101	-0.045	-0.235	-0.072
Mother - social assistance						
Childcare	-0.029	-0.037	-0.005	-0.012	0.023	0.016
Unemployment benefits	0.056	0.080	0.174*	0.184**	0.132	0.128*
Pension	-0.149**	-0.107*	-0.074	-0.057	0.012	0.037
Other	0.038	0.119***	-0.099	-0.059	-0.058	-0.001
Missing mother social						
assistance	0.131	0.016	0.096	-0.005	0.100	-0.003
Father - social assistance						
Childcare	0.048	0.026	0.005	-0.004	-0.015	-0.011
Unemployment benefits	0.174*	0.020	0.029	0.089	0.021	0.100
Pension	-0.106	-0.066	-0.074	-0.039	-0.101	-0.078
Other	-0.168*	-0.136	-0.121	-0.037	-0.146	-0.078
Missing fother social	-0.108	-0.150	-0.121	-0.087	-0.140	-0.078
assistance	-0.044	-0.016	-0.163*	-0.154*	-0.116	-0.089
Low weight at hirth	0 150***	0.005**	0 191***	0 115***	0 273***	0 231***
Nissing weight at birth	-0.139***	-0.095**	-0.181***	-0.113***	-0.273***	-0.231
Height for age soore	-0.320**	-0.230*	-0.318**	-0.299*	-0.192	-0.108
Forty shildhood hoalth issues	0.001	0.000***	0.001	0.000	0.001	0.000
Early childhood hearth issues	0.008	0.008	0.014	0.002	0.000	0.028
Speech Missing angela	-0.008	-0.008	0.014	0.002	0.066	0.028
Missing speech	-0.292	-0.177	-0.053	0.363	0.619	0.857
Hearing	0.057	0.014	-0.104	-0.068	-0.020	-0.104
Musculature	0.140	0.072	0.111	0.109	0.122	0.068
Circulatory	-0.081	0.045	0.037	0.107	0.013	0.062
Neurological	-0.229	-0.077	-0.488*	-0.243	-0.391*	-0.172
Gastrointestinal	0.095	0.196**	-0.062	-0.007	-0.209	-0.108
Asthma	0.027	0.063	-0.057	-0.021	-0.065	-0.046
Missing Asthma	-1.068***	-0.163	-0.513***	-0.100	-1.391***	-0.978***
Respiratory	-0.128*	-0.092	-0.034	0.014	-0.025	-0.002
Mental	0.087*	0.043	0.165***	0.126**	0.122**	0.059
Missing mental	1.033***	0.723***	1.021***	0.981***	0.9/2***	1.07/1***
Emaciation	-0.505***	-0.185	-0.270	0.062	-0.233	0.049
Obesity	-0.481*	-0.198	-0.357	-0.179	-0.180	0.021
Other	0.036	0.044	0.009	-0.017	-0.126	-0.120
Gender - Girl	0.543***	0.387***	0.408***	0.288***	-0.136***	-0.246***
Nursery School		-0.016		0.017		-0.006
Missing nursery		-0.110		0.327		-0.087
HOME -emotional subscale		0.005		-0.069***		-0.058***
HOME -cognitive subscale		0.073***		0.067***		0.045**
Missing HOME emotional		-4.705		69.334***		57.987***
Missing HOME cognitive		-72.780***		-67.042***		-45.609**
Internet		0.092***		0.087***		0.184***
Missing internet		0.148		-0.448		-0.264
Books at home - baseline >1000						
<50		-0.118*		-0.396***		-0.316***
~50		-0.037		-0.274***		-0.243***

50-150		-0.182***		-0.323***		-0.343***
150-300		-0.079 -0.205***			-0.155**	
300-600		-0.073*		-0.163***		-0.136**
600-1000		-0.072		-0.121**		-0.145***
Missing number of books		0.111		-0.075		0.108
Child own books (not textbooks)		0.221***		0.169***		0.221***
Missing child's books		0.089		0.113		0.026
Child's desk		0.150***		0.099**		0.085
Missing child's desk		-2.130***		-2.040***		-1.621***
Child own room		-0.022		-0.052*		-0.026
Child own computer		-0.015		-0.002		-0.034
Missing child's computer		1.264***		1.277***		1.390***
Education aspirations of parents - baseline elementary						
Secondary/vocational		0.359***		0.325***		0.261***
Higher		1.165***		0.944***		0.893***
Missing aspirations		0.845***		0.697***		0.626***
Abused before age 15		-0.049		0.028		-0.012
Missing abused		-0.125		-0.025		0.029
Abused after age 15		-0.062*		0.009		-0.012
Missing abused after age 15		0.054		0.052		0.048
Working for family age 15		-0.062		-0.006		0.003
Missing working for family age 15		0.107		0.411		0.052
Working for family primary school age		-0.077		-0.104**		-0.058
Missing working for family primary school age		-0.466***		-0.037		-0.226
R-squared	0.336	0.510	0.313	0.455	0.293	0.422

*** p<0.01, ** p<0.05, * p<0.1.Clustered standard errors at school level in brackets. Weighted by sampling weights.

Table A 16. Estimates for the coefficient on Roma using the main sample. Specification in columns with (1) control for socio-economic status and early childhood health while columns with (2) control in addition for home environment and school fixed effects

	(1)	(2)		(1)	(2)
Socio-emotional Skills	0.058	0.049	Expectations	-0.086	-0.097
	(0.059)	(0.058)	Age 17	(0.063)	(0.062)
Life Satisfaction	0.158**	8** 0.141* Expectations		-0.223***	-0.250***
Age 15	(0.074)	(0.073)	Age 21	(0.065)	(0.057)
Life Satisfaction	tisfaction -0.115*** -0.129*** Depression Age	Depression Age	-0.088	-0.101	
Age 21	(0.033)	(0.033)	15	(0.065)	(0.066)
Self Esteem Age	0.085	0.062	Learning Age	-0.064	-0.065
15	(0.067)	(0.068)	16	(0.072)	(0.071)

Self Esteem Age 18	0.080	0.042	Academic	-0.092	-0.082*
	(0.080)	(0.075)	Achievement Age 15	(0.062)	(0.048)
External Locus	0.045	0.064	Math Score	-0.308***	-0.256***
of Control Age 15	$\begin{array}{c} \text{Control Age} \\ 15 & (0.064) & (0.063) & \text{Age 15} \end{array}$	(0.056)	(0.054)		
External Locus	-0.018	0.019	Reading Score	-0.105*	-0.064
of Control Age 16	(0.075)	(0.078)	Age 15	(0.062)	(0.062)
Social	-0.035	-0.008	Health Age 15 -	0.113***	0.086**
Age 15	(0.084)	(0.089)	excellent	(0.033)	(0.034)
Social	0.090	0.057			
Competences Age 17	(0.073)	(0.072)			

*** p<0.01, ** p<0.05, * p<0.1.Clustered standard errors at school level in brackets. Weighted by sampling weights.

First Stage	Socio- emotional skills	Academic Achieveme nt	Reading Score	Math Score	Health excellent age 15
~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~	0.510**	0.487**	0.628**	0.388	0
	-0.251	-0.233	-0.263	-0.271	-0.124
-0.095	0.094	-0.051	-0.047	-0.274***	0.110***
-0.412***	0.072	-0.454***	-0.295**	-0.355***	-0.160**
-0.214***	0.051	-0.469***	-0.385***	-0.411***	-0.084**
-0.133***	0.062	-0.196***	-0.172***	-0.146**	-0.035
-0.167	0.039	-0.892***	-0.747***	-1.133***	0.145
-0.094	0.028	-0.455***	-0.504***	-0.563***	-0.033
-0.092	-0.025	-0.331***	-0.374***	-0.488***	-0.04
0	-0.05	-0.172***	-0.236***	-0.302***	-0.028
0.298	0.072	-0.21	0.122	-0.009	-0.11
-0.016	-0.017	0.022	0.034	0.072	-0.033
-0.243	0.139	0.302	0.510*	1.001***	-0.068
0.043	-0.021	0.004	0.013	0.026	-0.011
-0.281	-0.266	-0.055	-0.1	-0.27	-0.028
0.047	0.02	-0.049	-0.085*	-0.075*	0.089***
-0.013	0.051	-0.116	-0.217	-0.067	0.142**
0.038	0.005	-0.067	-0.008	-0.092*	0.141***
0.04	-0.009	-0.038	-0.122	-0.195	-0.085
-0.011	0.045	0.051	0.052	0.015	0.027
10.915	-44.51	-50.352	-51.073	-14.965	-26.343
-0.17	0.023	0.163	0.011	0.193*	-0.049
0.155*	-0.122	-0.06	-0.046	-0.082	-0.011
-0.015	-0.551	0.284	0.629	0.996***	-0.263
0.001	0.083**	0.039	0.005	0.02	-0.025
-0.394	0.832	0.641	0.496	0.401	-0.24
0.289***	-0.098	0.036	0.053	0.089	0.026
0.464	-0.039	-1.135**	-1.124***	-1.003***	0.366**
0	0.001	0.002	0.002	0.004**	0.002
0.061	-1.081	-1.708	-1.775	-3.863**	-1.804
0	0	0	0	0	0
-0.042	0.037	-0.102***	-0.033	-0.095**	0.008
0.016	-0.033	-0.122***	-0.107**	-0.098**	-0.035*
-0.039	0.05	-0.001	0.034	0.044	-0.002
-0.01	-0.015	-0.004	-0.032	-0.093	0.005
0.142***	-0.007	0.008	-0.025	0.048	0.013
- 141.394***	7.028	-8.263	25.089	-47.825	-12.761
-0.033	0.00	-0.054**	-0.023	-0.027	0.008
-0.151*	-0.08	-0.049	-0.08	-0.009	-0.05 /
-0.034	-0.133**	-0.05	0.043	0.075	-0.007
-0.015	-0.090	0.059	-0.003	-0.103**	-0.034
0.067	0.432	0.11	0.024	-0.033	0.199
-0.222	-0.140	0.013	0.175	0.154	0.033
	First Stage -0.095 -0.412*** -0.214*** -0.133*** -0.167 -0.094 -0.092 0 0.298 -0.016 -0.243 0.043 -0.281 0.047 -0.013 0.038 0.047 -0.013 0.38 0.047 -0.015 0.001 -0.394 0.289*** 0.464 0 0.061 0 -0.016 -0.033 -0.015 0.001 -0.394 0.289*** 0.464 0 0.016 -0.033 -0.015 0.033 -0.151* -0.033 -0.015 0.087 0.062 -0.222	Socio- emotional skills 0.510** -0.251 -0.095 0.094 -0.214*** 0.051 -0.133*** 0.062 -0.167 0.039 -0.094 -0.092 -0.092 -0.050 0.072 -0.094 -0.025 0 -0.092 -0.016 -0.017 -0.243 0.139 0.043 -0.261 -0.281 -0.263 0.043 -0.051 0.043 -0.051 0.043 -0.051 0.044 -0.038 0.005 0.044 -0.051 0.051 0.051 0.051 0.051 0.051 0.051 0.051 0.051 0.05	Socio- emotional skillsAcademic Achieveme nt 0.510^{**} 0.487^{**} -0.251 -0.233 -0.095 0.094 -0.051 -0.412^{***} 0.072 -0.454^{***} -0.214^{***} 0.051 -0.469^{***} -0.133^{***} 0.062 -0.196^{***} -0.167 0.039 -0.892^{***} -0.094 0.028 -0.455^{***} -0.092 -0.055 -0.31^{***} 0 -0.05 -0.172^{***} 0.298 0.072 -0.21 -0.016 -0.017 0.022 -0.243 0.139 0.302 0.043 -0.021 0.004 -0.281 -0.266 -0.055 0.047 0.02 -0.049 -0.013 0.051 -0.116 0.038 0.005 -0.067 0.04 -0.009 -0.38 -0.011 0.045 0.051 10.915 -44.51 -50.352 -0.17 0.023 0.163 0.155^{**} -0.122 -0.06 -0.015 -0.551 0.284 0.001 0.083^{**} 0.036 0.464 -0.039 -1.135^{**} 0 0.001 0.002 0.061 -1.081 -1.708 0 0.015 -0.004 -0.016 -0.033 -0.122^{***} -0.039 0.037 -0.004^{**} 0.001 0.001 -0.002^{***} 0.001 <t< td=""><td>Socio- emotional skillsAcademic Achieveme ntReading Score nt$0.510^{**}$$0.487^{**}$$0.628^{**}$$-0.251$$-0.233$$-0.263$$-0.095$$0.094$$-0.051$$-0.47^{**}$$-0.412^{***}$$0.072$$-0.454^{***}$$-0.295^{**}$$-0.214^{***}$$0.072$$-0.459^{***}$$-0.385^{***}$$-0.133^{***}$$0.062$$-0.166^{***}$$-0.172^{***}$$-0.167$$0.039$$-0.892^{***}$$-0.747^{***}$$-0.094$$0.028$$-0.455^{***}$$-0.504^{***}$$-0.092$$-0.025$$-0.331^{***}$$-0.747^{***}$$0$$-0.05$$-0.172^{***}$$-0.236^{***}$$0.092$$-0.025$$-0.331^{***}$$-0.374^{***}$$0$$-0.05$$-0.172^{***}$$-0.236^{***}$$0.298$$0.072$$-0.21$$0.122$$-0.016$$-0.017$$0.022$$0.034$$-0.281$$-0.266$$-0.055$$-0.1$$0.047$$0.02$$-0.049$$-0.085^{*}$$-0.013$$0.051$$-0.052$$-0.013$$0.047$$0.02$$-0.049$$-0.085^{*}$$-0.013$$0.051$$-0.162^{*}$$-0.008$$0.04$$-0.009$$-0.038$$-0.122^{*}$$-0.011$$0.045$$0.051$$0.052$$10.915$$-44.51$$-50.352$$-51.073$$-0.17$$0.023$$0.163$$0.011$$0.155^{*}$$-0.122^{*}$$-0.066$<td< td=""><td>First StageSocio- emotional skillsAcademic Achieveme ntReading ScoreMath Score$0.510^{**}$$0.487^{**}$$0.628^{**}$$0.388$$-0.251$$-0.233$$-0.263$$-0.271$$-0.095$$0.094$$-0.051$$-0.047$$-0.274^{***}$$-0.132^{***}$$0.051$$-0.469^{***}$$-0.355^{***}$$-0.355^{***}$$-0.133^{***}$$0.062$$-0.196^{***}$$-0.172^{***}$$-0.146^{**}$$-0.167$$0.039$$-0.892^{***}$$-0.747^{***}$$-1.133^{***}$$-0.094$$0.028$$-0.455^{***}$$-0.504^{***}$$-0.488^{***}$$0$$-0.05$$-0.172^{***}$$-0.314^{***}$$-0.488^{***}$$0$$-0.05$$-0.172^{***}$$-0.314^{***}$$-0.326^{***}$$-0.092$$-0.025$$-0.31^{***}$$-0.326^{***}$$-0.302^{***}$$0.092$$-0.025$$-0.31^{***}$$-0.302^{***}$$-0.326^{***}$$0.092$$-0.025$$-0.31^{***}$$-0.326^{***}$$-0.302^{***}$$0.172^{***}$$-0.210^{*}$$0.122^{*}$$-0.009^{*}$$-0.016$$-0.017$$0.022$$0.034$$0.072^{*}$$-0.281$$-0.266$$-0.055^{*}$$-0.1^{*}$$-0.27^{*}$$0.043$$-0.021$$-0.049^{*}$$-0.085^{*}$$-0.075^{*}$$-0.013$$0.051$$-0.052^{*}$$-0.11^{*}$$-0.067^{*}$$-0.033$$-0.023^{*}$$-0.049^{*}$$-0.052^{*}$$-0.014$$-0.0$</td></td<></td></t<>	Socio- emotional skillsAcademic Achieveme ntReading Score nt 0.510^{**} 0.487^{**} 0.628^{**} -0.251 -0.233 -0.263 -0.095 0.094 -0.051 -0.47^{**} -0.412^{***} 0.072 -0.454^{***} -0.295^{**} -0.214^{***} 0.072 -0.459^{***} -0.385^{***} -0.133^{***} 0.062 -0.166^{***} -0.172^{***} -0.167 0.039 -0.892^{***} -0.747^{***} -0.094 0.028 -0.455^{***} -0.504^{***} -0.092 -0.025 -0.331^{***} -0.747^{***} 0 -0.05 -0.172^{***} -0.236^{***} 0.092 -0.025 -0.331^{***} -0.374^{***} 0 -0.05 -0.172^{***} -0.236^{***} 0.298 0.072 -0.21 0.122 -0.016 -0.017 0.022 0.034 -0.281 -0.266 -0.055 -0.1 0.047 0.02 -0.049 -0.085^{*} -0.013 0.051 -0.052 -0.013 0.047 0.02 -0.049 -0.085^{*} -0.013 0.051 -0.162^{*} -0.008 0.04 -0.009 -0.038 -0.122^{*} -0.011 0.045 0.051 0.052 10.915 -44.51 -50.352 -51.073 -0.17 0.023 0.163 0.011 0.155^{*} -0.122^{*} -0.066 <td< td=""><td>First StageSocio- emotional skillsAcademic Achieveme ntReading ScoreMath Score$0.510^{**}$$0.487^{**}$$0.628^{**}$$0.388$$-0.251$$-0.233$$-0.263$$-0.271$$-0.095$$0.094$$-0.051$$-0.047$$-0.274^{***}$$-0.132^{***}$$0.051$$-0.469^{***}$$-0.355^{***}$$-0.355^{***}$$-0.133^{***}$$0.062$$-0.196^{***}$$-0.172^{***}$$-0.146^{**}$$-0.167$$0.039$$-0.892^{***}$$-0.747^{***}$$-1.133^{***}$$-0.094$$0.028$$-0.455^{***}$$-0.504^{***}$$-0.488^{***}$$0$$-0.05$$-0.172^{***}$$-0.314^{***}$$-0.488^{***}$$0$$-0.05$$-0.172^{***}$$-0.314^{***}$$-0.326^{***}$$-0.092$$-0.025$$-0.31^{***}$$-0.326^{***}$$-0.302^{***}$$0.092$$-0.025$$-0.31^{***}$$-0.302^{***}$$-0.326^{***}$$0.092$$-0.025$$-0.31^{***}$$-0.326^{***}$$-0.302^{***}$$0.172^{***}$$-0.210^{*}$$0.122^{*}$$-0.009^{*}$$-0.016$$-0.017$$0.022$$0.034$$0.072^{*}$$-0.281$$-0.266$$-0.055^{*}$$-0.1^{*}$$-0.27^{*}$$0.043$$-0.021$$-0.049^{*}$$-0.085^{*}$$-0.075^{*}$$-0.013$$0.051$$-0.052^{*}$$-0.11^{*}$$-0.067^{*}$$-0.033$$-0.023^{*}$$-0.049^{*}$$-0.052^{*}$$-0.014$$-0.0$</td></td<>	First StageSocio- emotional skillsAcademic Achieveme ntReading ScoreMath Score 0.510^{**} 0.487^{**} 0.628^{**} 0.388 -0.251 -0.233 -0.263 -0.271 -0.095 0.094 -0.051 -0.047 -0.274^{***} -0.132^{***} 0.051 -0.469^{***} -0.355^{***} -0.355^{***} -0.133^{***} 0.062 -0.196^{***} -0.172^{***} -0.146^{**} -0.167 0.039 -0.892^{***} -0.747^{***} -1.133^{***} -0.094 0.028 -0.455^{***} -0.504^{***} -0.488^{***} 0 -0.05 -0.172^{***} -0.314^{***} -0.488^{***} 0 -0.05 -0.172^{***} -0.314^{***} -0.326^{***} -0.092 -0.025 -0.31^{***} -0.326^{***} -0.302^{***} 0.092 -0.025 -0.31^{***} -0.302^{***} -0.326^{***} 0.092 -0.025 -0.31^{***} -0.326^{***} -0.302^{***} 0.172^{***} -0.210^{*} 0.122^{*} -0.009^{*} -0.016 -0.017 0.022 0.034 0.072^{*} -0.281 -0.266 -0.055^{*} -0.1^{*} -0.27^{*} 0.043 -0.021 -0.049^{*} -0.085^{*} -0.075^{*} -0.013 0.051 -0.052^{*} -0.11^{*} -0.067^{*} -0.033 -0.023^{*} -0.049^{*} -0.052^{*} -0.014 -0.0

Table A 17. IV regression estimates on the main sample

Discounted meals at school	-0.017	-0.023	0.007	-0.011	0.015	0.005
Missing discounted meals	0.216	-0.365	-0.063	-0.07	0.161	-0.222
Free textbooks	0.055	-0.094**	-0.054	-0.074*	-0.047	-0.051**
Missing free textbooks	0.386*	-0.122	-0.168	-0.37	-0.245	-0.16
Education aid	-0.013	0.106**	-0.009	0.080*	0.018	0.043*
Missing education aid	-0.142	-0.003	-0.011	-0.017	-0.168	-0.012
Mother - social assistance						
Childcare	0.006	-0.069	-0.032	-0.009	0.018	-0.017
Unemployment benefits	0.024	0.109	0.042	0.155	0.125	-0.007
Pension	-0.026	-0.051	-0.132*	-0.049	0.029	-0.013
Other	0.009	0.044	0.038	-0.098	-0.061	0.008
Missing mother social	0.017	-0.061	0.125	0.088	0.088	-0.055
assistance	0.017	-0.001	0.125	0.088	0.088	-0.055
Father - social assistance						
Childcare	-0.031	-0.054	0.064	0.026	-0.002	-0.005
Unemployment benefits	0.031	0.088	0.162	0.012	0.013	0.061
Pension	-0.001	0.027	-0.102	-0.068	-0.095	0.009
Other	-0.097	-0.147	-0.121	-0.056	-0.103	0.01
Missing father social assistance	-0.038	0.199**	-0.029	-0.142	-0.091	-0.017
Low weight at birth	0.003	-0.056	-0.158***	-0.180***	-0.275***	-0.026
Missing weight at birt	-0.153	0.06	-0.231	-0.186	-0.1	0.002
Height for age score	-0.001**	0.001**	0.001***	0.001*	0.001*	0
Early childhood health issues						
Speech	0.098**	-0.169***	-0.046	-0.039	0.038	-0.042**
Missing speech	-0.657	0.633**	-0.009	0.344	0.86	-0.356***
Hearing	0.084	-0.380***	0.024	-0.15	-0.07	-0.102*
Musculature	0.041	0.008	0.123	0.086	0.093	-0.075*
Circulatory	0.145*	-0.144	-0.137	-0.041	-0.038	-0.154***
Neurological	-0.018	-0.447	-0.21	-0.461**	-0.398**	-0.11
Gastrointestinal	-0.147	0.242*	0.153	0.019	-0.164	0.103
Asthma	0.045	-0.095	0.01	-0.082	-0.072	-0.066
Missing Asthma	0.088	-0.041	-1.042***	-0.476***	-1.359***	-0.096
Respiratory	-0.031	-0.152*	-0.115	-0.016	-0.024	-0.056
Mental	0.109**	-0.038	0.042	0.103	0.085	-0.072**
Missing mental	1.060***	-0.452	0.589**	0.399	0.602*	-0.397***
Emaciation	-0.339	-0.513*	-0.349	-0.052	-0.102	-0.098
Obesity	0.167	-0.714*	-0.547*	-0.448	-0.256	-0.106
Other	-0.104	-0.483	0.078	0.067	-0.089	-0.210*
Gender - Girl	0.054*	-0.201***	0.519***	0.375***	-0.155***	-0.063***
Young children in HH when	0.046*					
child was 6 years old	(-0.024)					
	(-0.024)					
Separated from mother age 0-7	-0.554***					
	(-0.129)					
Separated from father age 0-7	-0.077					
	(-0.059)					
Stepfather age 0-7	0.198					
	-0.125					
Stepmother age 0-7	-0.051					
	-0.204					

*** p<0.01, ** p<0.05, * p<0.1.Clustered standard errors at school level in brackets. Weighted by sampling weights.

	First Stage	Socio- emotional skills	Academic Achievemen t	Reading Score	Math Score	Health excellent age 15
			0.455*		0.050	0.012
EC Investment		0.644**	0.477*	0.762**	0.378	-0.013
		-0.323	-0.271	-0.309	-0.308	-0.13
Roma	-0.158***	0.150**	0.042	-0.029	-0.193***	0.062**
Mother education - baseline higher education						
0-8 grade	-0.364***	0.138	-0.490***	-0.290**	-0.424***	-0.116**
Vocational	-0.217***	0.103	-0.477***	-0.355***	-0.412***	-0.051
Secondary	-0.112***	0.059	-0.209***	-0.169***	-0.187***	-0.042
Missing mother education	-0.137	-0.004	-0.633***	-0.632***	-0.704***	-0.002
Father education- baseline higher education						
0-8 grade	-0.138**	-0.04	-0.427***	-0.458***	-0.579***	-0.051*
Vocational	-0.06	-0.06	-0.353***	-0.392***	-0.496***	-0.050**
Secondary	-0.001	-0.064	-0.154***	-0.242***	-0.279***	-0.009
Missing father education	0.09	0.03	-0.331**	-0.125	-0.218	-0.114
Mother employed	0.023	-0.05	0.024	-0.017	0.039	-0.015
Missing mother employed	-0.166	-0.196	-0.038	0.298	0.217	-0.002
Father employed	0.101*	-0.088	0.004	-0.021	-0.005	-0.023
Missing father employed	0.017	-0.426**	0.032	-0.082	-0.146	-0.083
Father excellent health	0.05	0.033	-0.042	-0.088**	-0.047	0.142***
Missing father health	-0.093	0.16	-0.029	0.022	0.028	0.121*
Mother excellent health	0.029	-0.03	-0.026	-0.001	-0.062*	0.152***
Missing mother health	0.063	0.038	-0.11	-0.196*	-0.105	0.002
Logarithm of income	-0.023	0.028	0.075**	0.099**	0.081**	0.012
Missing income	22.733	-27.141	-74.274**	-97.450**	-80.260**	-12.05
Logarithm of HH size	-0.055	0.031	0.001	-0.118	-0.013	-0.025
Bathroom in the house	0.125	-0.027	-0.036	-0.037	-0.02	-0.012
Missing bathroom	0.15	0.101	0.407	0.181	0.419	-0.245
Car	0.01	0.049	0.053**	0.012	0.054**	-0.017
Missing car	0.077	0.19	0.216	0.049	0.253	-0.013
Computer	0.228***	-0.069	0.058	0.061	0.095	0.023
Missing computer	0.171	-0.166	-0.780***	-0.395	-0.617***	0.154
Apartment size/person	-0.001	0.001	0.002	0.002	0.002*	0.002**
Missing apartment size	0.657	-1.266	-1.76	-1.468	-2.130*	-1.994**
Rooms/person	0	0	0	0	0	0
People in HH with:	0.04111	0.017	0.072	0.05.	0.070	0.007
Regular work	-0.064**	0.048	-0.072**	-0.004	-0.078***	-0.002
Unemployed	-0.044	-0.018	-0.076***	-0.060*	-0.057	-0.026*
Old age	-0.031	0.011	-0.034	0.013	-0.029	-0.036
Pensioners	0.019	0.013	0.03	0.001	-0.017	0.038**
Neighborhood index	0.140^{***}	-0.009	0.007	-0.035	0.024	0.026
Missing neighborhood index	-139.750***	8.965	-7.426	34.479	-24.176	-25.63

Table A18. Robustness check. IV regression results on the full available sample

Siblings	-0.031**	0.015	-0.007	0.008	-0.008	0.006
Poverty - Food	-0.115**	-0.059	-0.027	-0.017	-0.045	-0.049
Poverty - Rent	-0.054	-0.073	-0.087**	0.007	0.007	-0.029*
Poverty - Heating	-0.004	-0.085	0.035	0.025	-0.05	0.018
Missing poverty	0.003	0.195	0.035	-0.063	0.158	0.085
Free meals at schools	0.009	-0.083	-0.028	-0.123**	-0.054	0.012
Missing free meals	-0 373*	-0.09	0.223	0.445*	0.311*	0.139
Discounted meals at	0.02	0.005	0.011	0.012	0.02	0.000
school	-0.03	-0.025	0.011	0.013	0.03	-0.002
Missing discounted meals	0.131	-0.249	0.037	0.069	0.168	-0.153
Free textbooks	0.072**	-0.102***	-0.105***	-0.124***	-0.079**	-0.02
Missing free textbooks	0.348*	-0.323	-0.301	-0.748**	-0.155	0.253
Education aid	-0.018	0.094***	0.018	0.066	0.018	0.040**
Missing education aid	-0.09	-0.011	-0.1	0.176	0.022	0.025
Mother - social						
Children	0.000	0.052	0.007	0.025	0.010	0.017
Unemployment	0.009	-0.053	0.007	0.025	0.019	-0.016
benefits	0.077	0.052	0.052	0.105	0.11	0.018
Pension	-0.038	-0.071	-0.07	-0.024	-0.048	-0.055**
Other	-0.036	0.004	0.04	-0.057	-0.059	0.012
Missing mother social	0.011	0.110	0 140 **	0.17644	0.1.6.4*	0.070**
assistance	-0.011	0.112	0.148**	0.176**	0.164*	-0.072**
Father - social						
assistance						
Childcare	-0.039	-0.047	0.079**	0.037	0.031	-0.011
Unemployment	0.053	-0.043	-0.105	-0.111	-0.094	-0.008
Pension	-0.046	-0.081	-0.069	-0.051	-0.064	-0.036
Other	-0.021	0.165**	-0.034	-0.175**	-0.122**	0.02
Missing father social	01021	01100	01001	01170	01122	0.02
assistance	-0.02	-0.071	-0.142***	-0.162***	-0.255***	-0.011
Low weight at birth	-0.020	-0.071	-0.142***	-0.162***	-0.255***	-0.011
Missing weight at birth	-0.419***	0.094	0.028	0.062	0.018	0.041
Height for age score	-0.001**	0.001***	0.001*	0.001**	0.001	-0.000
EC health issues						
Speech	0.094***	-0.183***	-0.053	-0.058	0.031	-0.050***
Missing speech	-0.735**	0.717*	-0.023	0.197	0.231	-0.166
Hearing	0.084	-0.305***	-0.044	-0.160	-0.104	-0.110***
Musculature	0.011	0.064	0.095	0.060	0.119	-0.091***
Circulatory	0.017	-0.105	-0.107	-0.055	-0.076	-0.134***
Neurological	-0.310**	-0.236	0.050	-0.244	-0.340*	-0.080
Gastrointestinal	-0.008	-0.236	0.127	-0.036	-0.129	-0.024
Asthma	0.019	-0.066	0.035	-0.017	0.022	-0.074**
Missing Asthma	0.029	0.010	-1.038***	-0.545***	-1.405***	-0.078
Respiratory	0.061	-0.191***	-0.069	-0.023	-0.026	-0.096***
Mental	0.088**	-0.009	0.036	0.044	0.033	-0.052**
Missing mental	1.075***	-0.619*	0.666**	0.354	0.693**	-0.342**
Emaciation	-0.303	-0.846***	-0.544***	-0.121	-0.289*	-0.138
Obesity	-0.206	0.014	-0.400**	-0.374	-0.343	-0.031
Other	-0.062	-0.897***	-0.279	-0.134	-0.222	-0.243***
Gender - Girl	0.052***	-0.246***	0.519***	0.366***	-0.163***	-0.071***

Younger children in HH when child was 6	0.016						
years old	-0.016						
Separated from mother	-0.238***						
age 0-7	-0.089						
Separated from father	-0.068*						
age 0-7	-0.035						
Stanfathan and 0.7	0.137*						
Stepratier age 0-7	-0.077						
	0.073						
Stepmotner age 0-7	-0.143						
Observations	9,902	9,773	9,637	9,902	9,083	9,821	

*** p<0.01, ** p<0.05, * p<0.1.Clustered standard errors at school level in brackets. Weighted by sampling weights.